STIC-Biotech/ChemLib 173379 GRPE

From: Russel, Jeffrey
Sent: Friday, December 02, 2005 3:56 PM
To: STIC-Biotech/ChemLib
Subject: Database Search Request

Requester:
Jeffrey Russel (TC1600)
Art Unit:
1654
Employee Number:
62785
Office Location:
REM 3D19
Phone_Number:
571-272-0969
Mailbox Number:
REM 3C18

Checked JRL 12-21-2005

Case serial number:
10/772,164
Class / Subclass(es):
NA
Earliest Priority Filing Date:
10/21/1997
Format preferred for results:
Diskette
aa 17
Search Topic Information:
Please search SEQ ID NO:1 in the U.S. patent application sequence database (pending, published, and issued), and in Geneseq-Uniprot-PIR.

Please search the partial sequence CCXXCC (residues 9-14 of SEQ ID NO:1) in STN. If there are many hits, please require X to be R, E, A, L, or M.

Thank you.
Special Instructions and Other Comments:

MEJ

****
Searcher:_____
Searcher Phone:_____
Date Searcher Picked up:_____
Date completed:_____
Searcher Prep Time:_____
Online Time:_____

****
Type of Search
NA#_____ AA#:_____
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure #:_____Text:_____
Inventor:_____ Litigation:_____

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other (Specify):_____

Date completed: 12.12
Searcher: Beverly e2528
Terminal time: _____
Elapsed time: _____
CPU time: _____
Total time: _____
Number of Searches: _____
Number of Databases: _____

Search Site
_____ STIC
_____ CM-1
_____ Pre-S

Type of Search
_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors
_____ IG
✓ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
✓ Other CAS

1590 (5-90)

FILE 'REGISTRY' ENTERED AT 12:36:37 ON 12 DEC 2005

Property values tagged with IC are from the ZIC/VINITI data file provided by InfoChem.

STRUCTURE FILE UPDATES: 11 DEC 2005 HIGHEST RN 869700-38-9
DICTIONARY FILE UPDATES: 11 DEC 2005 HIGHEST RN 869700-38-9

New CAS Information Use Policies, enter HELP USAGETERMS for details.

TSCA INFORMATION NOW CURRENT THROUGH JULY 14, 2005

Please note that search-term pricing does apply when conducting SmartSELECT searches.

****
*
* The CA roles and document type information have been removed from *
* the IDE default display format and the ED field has been added, *
* effective March 20, 2005. A new display format, IDERL, is now *
* available and contains the CA role and document type information. *
*
****

Structure search iteration limits have been increased. See HELP SLIMITS for details.

REGISTRY includes numerically searchable data for experimental and predicted properties as well as tags indicating availability of experimental property data in the original document. For information on property searching in REGISTRY, refer to:

http://www.cas.org/ONLINE/UG/regprops.html

L1 7 S CCRECC/SQSP

L1 ANSWER 1 OF 7 REGISTRY COPYRIGHT 2005 ACS on STN
RN 849178-34-3 REGISTRY
CN L-Cysteine, L-cysteinyl-L-cysteinyl-L-arginyl-L-α-glutamyl-L-cysteinyl- (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 43: PN: WO2005028615 SEQID: 58 unclaimed sequence
SQL 6

SEQ 1 CCRECC
======
HITS AT: 1-6

RELATED SEQUENCES AVAILABLE WITH SEQLINK

REFERENCE 1: 142:368740

L1 ANSWER 2 OF 7 REGISTRY COPYRIGHT 2005 ACS on STN
RN 600706-61-4 REGISTRY
CN L-Isoleucine, L-arginyl-L-valyl-L-α-aspartyl-L-alanyl-L-alanyl-L-alanyl-L-arginyl-L-α-glutamyl-L-alanyl-L-cysteinyl-L-cysteinyl-L-

arginyl-L-α-glutamyl-L-cysteinyl-L-cysteinyl-L-alanyl-L-threonyl-L-alanyl- (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 16: PN: WO03075856 SEQID: 16 unclaimed sequence
SQL 19

SEQ 1 RVDAAAREAC CRECCATAI
= =====
HITS AT: 10-15

REFERENCE 1: 139:256227

L1 ANSWER 3 OF 7 REGISTRY COPYRIGHT 2005 ACS on STN
RN 439806-84-5 REGISTRY
CN L-Cysteine, L-cysteinyl-L-cysteinyl-L-arginyl-L-α-glutamyl-L-cysteinyl-, cyclic 1,2:5,6-[(3',6'-dihydroxy-3-oxospiro[isobenzofuran-1(3H),9'-[9H]xanthene]-4',5'-diyl)bis[arsonodithioite]] (9CI) (CA INDEX NAME)
SQL 6

SEQ 1 CCRECC
======
HITS AT: 1-6

RELATED SEQUENCES AVAILABLE WITH SEQLINK

REFERENCE 1: 137:59787

L1 ANSWER 4 OF 7 REGISTRY COPYRIGHT 2005 ACS on STN
RN 394709-23-0 REGISTRY
CN L-Isoleucine, L-arginyl-L-valyl-L-α-aspartyl-L-alanyl-L-alanyl-L-alanyl-L-arginyl-L-α-glutamyl-L-alanyl-L-cysteinyl-L-cysteinyl-L-arginyl-L-α-glutamyl-L-cysteinyl-L-cysteinyl-L-alanyl-L-arginyl-L-alanyl- (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 16: PN: WO0210364 SEQID: 16 unclaimed sequence
CN 59: PN: WO03075856 FIGURE: 4 unclaimed sequence
SQL 19

SEQ 1 RVDAAAREAC CRECCARAI
= =====
HITS AT: 10-15

REFERENCE 1: 139:256227

REFERENCE 2: 136:146127

L1 ANSWER 5 OF 7 REGISTRY COPYRIGHT 2005 ACS on STN
RN 268741-28-2 REGISTRY
CN L-Alanine, L-tryptophyl-L-α-glutamyl-L-alanyl-L-alanyl-L-alanyl-L-arginyl-L-α-glutamyl-L-alanyl-L-cysteinyl-L-cysteinyl-L-arginyl-L-α-glutamyl-L-cysteinyl-L-cysteinyl-L-alanyl-L-arginyl- (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 46: PN: WO0114578 PAGE: 38 unclaimed sequence
CN 4: PN: WO0047220 SEQID: 48 unclaimed sequence
CN 8: PN: US20040014071 SEQID: 4 unclaimed sequence
CN 8: PN: WO2005038029 SEQID: 8 unclaimed sequence
CN 9: PN: WO0153325 PAGE: 32 claimed sequence

SQL 17

SEQ 1 WEAAAREACC RECCARA
== ====
HITS AT: 9-14

RELATED SEQUENCES AVAILABLE WITH SEQLINK

REFERENCE 1: 142:425896

REFERENCE 2: 140:124793

REFERENCE 3: 135:149588

REFERENCE 4: 134:204756

REFERENCE 5: 133:172215

REFERENCE 6: 132:344976

L1 ANSWER 6 OF 7 REGISTRY COPYRIGHT 2005 ACS on STN
RN 223673-79-8 REGISTRY
CN L-Alanine, L-alanyl-L-α-glutamyl-L-alanyl-L-alanyl-L-alanyl-L-arginyl-L-α-glutamyl-L-alanyl-L-cysteinyl-L-cysteinyl-L-arginyl-L-α-glutamyl-L-cysteinyl-L-cysteinyl-L-alanyl-L-arginyl- (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 5: PN: WO0047220 SEQID: 49 unclaimed sequence
SQL 17

SEQ 1 AEAAAREACC RECCARA
== ====
HITS AT: 9-14

REFERENCE 1: 137:59787

REFERENCE 2: 133:172215

REFERENCE 3: 130:308804

L1 ANSWER 7 OF 7 REGISTRY COPYRIGHT 2005 ACS on STN
RN 223673-78-7 REGISTRY
CN L-Alaninamide, N-acetyl-L-tryptophyl-L-α-glutamyl-L-alanyl-L-alanyl-L-alanyl-L-arginyl-L-α-glutamyl-L-alanyl-L-cysteinyl-L-cysteinyl-L-arginyl-L-α-glutamyl-L-cysteinyl-L-cysteinyl-L-alanyl-L-arginyl- (9CI) (CA INDEX NAME)
SQL 17

SEQ 1 WEAAAREACC RECCARA
== ====
HITS AT: 9-14

RELATED SEQUENCES AVAILABLE WITH SEQLINK

REFERENCE 1: 137:59787

REFERENCE 2: 130:308804

FILE 'CAPLUS' ENTERED AT 12:36:39 ON 12 DEC 2005

Searcher : Shears 571-272-2528

FILE COVERS 1907 - 12 Dec 2005 VOL 143 ISS 25
FILE LAST UPDATED: 11 Dec 2005 (20051211/ED)

Effective October 17, 2005, revised CAS Information Use Policies apply. They are available for your review at:

http://www.cas.org/infopolicy.html

L3 11 L1

L3 ANSWER 1 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 29 Apr 2005
ACCESSION NUMBER: 2005:371402 CAPLUS
DOCUMENT NUMBER: 142:425896
TITLE: Beetle luciferase reporter protein with various modification motif increase or decrease luminescence activity in the present or absent of exogenous agent
INVENTOR(S): Fan, Frank; Lewis, Martin Ken; Schultz, John W.; Wood, Keith V.; Butler, Braeden
PATENT ASSIGNEE(S): Promega Corporation, USA
SOURCE: PCT Int. Appl., 178 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2005038029	A2	20050428	WO 2004-US32705	20041001
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW				
RW: BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG				
US 2005153310	A1	20050714	US 2004-957433	20041001

PRIORITY APPLN. INFO.: US 2003-510187P P 20031010

AB The current invention provides beetle luciferase reporter protein with various modifications. The reporter protein with modified motif in the absence or the present of an exogenous agent may enhance or inhibit luciferase activity.

IT **268741-28-2**
RL: PRP (Properties)
(unclaimed sequence; beetle luciferase reporter protein with various modification motif increase or decrease luminescence activity in the present or absent of exogenous agent)

L3 ANSWER 2 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 01 Apr 2005
ACCESSION NUMBER: 2005:283572 CAPLUS
DOCUMENT NUMBER: 142:368740
TITLE: Plasmid vectors containing recombination sites and topoisomerase recognition sites for detecting promoter activity and expressing fusion proteins
INVENTOR(S): Welch, Peter J.; Chesnut, Jonathan D.; Bennett, Robert P.; Frimpong, Kenneth; Leong, Louis; Fan, James; Yim, Harry; Vozza-Brown, Laura
PATENT ASSIGNEE(S): Invitrogen Corporation, USA
SOURCE: PCT Int. Appl., 378 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2005028615	A2	20050331	WO 2004-US20747	20040628
WO 2005028615	A3	20050825		
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW				
RW: BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG				
US 2005095615	A1	20050505	US 2004-877952	20040628

PRIORITY APPLN. INFO.: US 2003-482504P P 20030626
US 2003-487301P P 20030716
US 2003-511634P P 20031017

AB The present invention provides nucleic acid mols. comprising one or more nucleic acid sequences encoding a polypeptide having a detectable activity, and in particular β-lactamase, said vectors comprising multiple recombination sites and/or topoisomerase recognition sites operably linked to a promoter. The present invention also provides methods of joining such nucleic acid mols. to nucleic acid mols. to be

assayed for promoter activity. The present invention also relates to methods of preparing fusion proteins comprising a polypeptide of interest and a polypeptide having a detectable activity. The GeneBLAzer System comprises the β-lactamase gene coupled with a fluorescence resonance energy transfer (FRET)-enabled substrate (CCF2, CCF2-FA, CCF2-AM, or CCF4-AM) and is an excellent reporter system for promoter studies in mammalian cells. A 'promoterless" β-lactamase vector (pGeneBlazer) may be constructed as a bidirectional TOPO vector, allowing PCR amplification of one or more promoters of interest and cloning of the promoters upstream of the β-lactamase gene. Recombination sites in combination with topoisomerase recognition sites allow joining of nucleic acids for expression of fusion proteins. Thus invention also uses nucleic acid regions encoding peptides with affinity for arsenic (Cys-Cys-X-X-Cys-Cys). The design, construction, and sequences of a variety of plasmid vectors is described.

IT **849178-34-3**
RL: PRP (Properties)
(unclaimed sequence; plasmid vectors containing recombination sites and topoisomerase recognition sites for detecting promoter activity and expressing fusion proteins)

L3 ANSWER 3 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 23 Jan 2004
ACCESSION NUMBER: 2004:59670 CAPLUS
DOCUMENT NUMBER: 140:124793
TITLE: Methods for the detection, analysis and isolation of nascent proteins using non-radioactive markers
INVENTOR(S): Rothschild, Kenneth J.; Gite, Sadanand; Olejnik, Jerzy; Lim, Mark
PATENT ASSIGNEE(S): Ambergen, Inc., USA
SOURCE: U.S. Pat. Appl. Publ., 147 pp., Cont.-in-part of U.S. Ser. No. 49,332.
CODEN: USXXCO
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 3
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 2004014071	A1	20040122	US 2003-339712	20030110
US 6306628	B1	20011023	US 1999-382736	19990825
WO 2001014578	A1	20010301	WO 2000-US23233	20000823
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
US 2005009013	A1	20050113	US 2001-813197	20010320
US 6875592	B2	20050405		
US 2003190643	A1	20031009	US 2002-264127	20021003
US 2005032078	A1	20050210	US 2003-719523	20031121
CA 2512552	AA	20040729	CA 2004-2512552	20040109

WO 2004063714 A2 20040729 WO 2004-US528 20040109
W: AE, AE, AG, AL, AL, AM, AM, AM, AT, AT, AU, AU, AZ, AZ, BA, BB, BG, BG, BR, BR, BW, BY, BY, BZ, BZ, CA, CH, CN, CN, CO, CO, CR, CR, CU, CU, CZ, CZ, DE, DE, DK, DK, DM, DZ, EC, EC, EE, EE, EG, ES, ES, FI, FI, GB, GD, GE, GE, GH, GH, GH, GM, HR, HR, HU, HU, ID, IL, IN, IS, JP, JP, KE, KE, KG, KG, KP, KP, KP, KR, KR, KZ, KZ, KZ, LC, LK, LR, LS, LS, LT, LU, LV, MA, MD, MD, MG, MK, MN, MW, MX, MX, MZ
EP 1581797 A2 20051005 EP 2004-701238 20040109
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR, BG, CZ, EE, HU, SK
PRIORITY APPLN. INFO.: US 1999-382736 A2 19990825
WO 2000-US23233 W 20000823
US 2002-49332 A2 20020621
US 1999-382950 A 19990825
US 2001-813197 A1 20010320
US 2003-339712 A 20030110
WO 2004-US528 W 20040109

AB This invention relates to non-radioactive markers that facilitate the detection and anal. of nascent proteins translated within cellular or cell-free translation systems. Nascent proteins containing these markers can be rapidly and efficiently detected, isolated and analyzed without the handling and disposal problems associated with radioactive reagents. Preferred markers are dipyrromethene boron difluoride (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene) dyes.

IT **268741-28-2**
RL: PRP (Properties)
(unclaimed sequence; methods for the detection, anal. and isolation of nascent proteins using non-radioactive markers)

L3 ANSWER 4 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 19 Sep 2003
ACCESSION NUMBER: 2003:737532 CAPLUS
DOCUMENT NUMBER: 139:256227
TITLE: Methods for enhancing oligonucleotide-directed nucleic acid sequence alteration using repair proteins, histone deacetylase inhibitors, λ phage β proteins and hydroxyurea for use in therapy of blood diseases
INVENTOR(S): Kmiec, Eric B.; Parekh-Olmedo, Hetal; Brachman, Erin E.
PATENT ASSIGNEE(S): University of Delaware, USA
SOURCE: PCT Int. Appl., 135 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003075856	A2	20030918	WO 2003-US7217	20030307

WO 2003075856	A3	20040624			
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW					
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG					
CA 2478479	AA	20030918	CA 2003-2478479		20030307
US 2003207451	A1	20031106	US 2003-384918		20030307
EP 1490013	A2	20041229	EP 2003-716412		20030307
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR, BG, CZ, EE, HU, SK					
JP 2005518817	T2	20050630	JP 2003-574132		20030307
PRIORITY APPLN. INFO.:			US 2002-363053P	P	20020307
			US 2002-363054P	P	20020307
			US 2002-363341P	P	20020307
			US 2002-416983P	P	20021007
			WO 2003-US7217	W	20030307

AB Improved methods, compns., and kits for oligonucleotide-mediated nucleic acid sequence alteration using repair proteins, histone deacetylase inhibitors and hydroxyurea are provided. These methods may be use for treatment of blood disorders.

IT **394709-23-0 600706-61-4**
RL: PRP (Properties)
(unclaimed sequence; methods for enhancing oligonucleotide-directed nucleic acid sequence alteration using repair proteins, histone deacetylase inhibitors, λ phage β proteins and hydroxyurea for use in therapy of blood diseases)

L3 ANSWER 5 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 02 May 2002

ACCESSION NUMBER:	2002:326835 CAPLUS
DOCUMENT NUMBER:	137:59787
TITLE:	New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications
AUTHOR(S):	Adams, Stephen R.; Campbell, Robert E.; Gross, Larry A.; Martin, Brent R.; Walkup, Grant K.; Yao, Yong; Llopis, Juan; Tsien, Roger Y.
CORPORATE SOURCE:	Department of Pharmacology, Department of Chemistry and Biochemistry, Howard Hughes Medical Institute and Biomedical Sciences Program, University of California San Diego, La Jolla, CA, 92093-0647, USA
SOURCE:	Journal of the American Chemical Society (2002), 124(21), 6063-6076 CODEN: JACSAT; ISSN: 0002-7863
PUBLISHER:	American Chemical Society
DOCUMENT TYPE:	Journal

LANGUAGE: English

AB We recently introduced a method (Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Science 1998, 281, 269-272 and Griffin, B. A.; Adams, S. R.; Jones, J.; Tsien, R. Y. Methods Enzymol. 2000, 327, 565-578) for site-specific fluorescent labeling of recombinant proteins in living cells. The sequence Cys-Cys-Xaa-Xaa-Cys-Cys, where Xaa is an noncysteine amino acid, is genetically fused to or inserted within the protein, where it can be specifically recognized by a membrane-permeant fluorescein derivative with two As(III) substituents, FlAsH, which fluoresces only after the arsenics bind to the cysteine thiols. We now report kinetics and dissociation consts. (.apprx.10-11 M) for FlAsH binding to model tetracysteine peptides. Affinities in vitro and detection limits in living cells are optimized with Xaa-Xaa = Pro-Gly, suggesting that the preferred peptide conformation is a hairpin rather than the previously proposed α-helix. Many analogs of FlAsH have been synthesized, including ReAsH, a resorufin derivative excitable at 590 nm and fluorescing in the red. Analogous biarsenicals enable affinity chromatog., fluorescence anisotropy measurements, and electron-microscopic localization of tetracysteine-tagged proteins.

IT **223673-78-7 223673-79-8 439806-84-5**
RL: ARU (Analytical role, unclassified); ANST (Analytical study)
(biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo)

REFERENCE COUNT: 45 THERE ARE 45 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L3 ANSWER 6 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 10 Feb 2002
ACCESSION NUMBER: 2002:107524 CAPLUS
DOCUMENT NUMBER: 136:146127
TITLE: Methods for enhancing targeted gene alteration in cells having altered activity of DNA repair proteins using chimeric RNA-DNA double-stranded hairpin oligonucleotides
INVENTOR(S): Kmiec, Eric B.; Gamper, Howard B.; Rice, Michael C.; Liu, Li
PATENT ASSIGNEE(S): University of Delaware, USA
SOURCE: PCT Int. Appl., 48 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 4
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2002010364	A2	20020207	WO 2001-US23770	20010727
WO 2002010364	A3	20030925		

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI,

	CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG				
CA 2417344		AA	20020207	CA 2001-2417344	20010727
EP 1364008		A2	20031126	EP 2001-957311	20010727
	R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR				
US 2003217377		A1	20031120	US 2002-209787	20020730
US 2003215947		A1	20031120	US 2003-351662	20030124
PRIORITY APPLN. INFO.:				US 2000-220999P	P 20000727
				US 2000-244989P	P 20001030
				US 2000-192176P	P 20000327
				US 2000-192179P	P 20000327
				US 2000-208538P	P 20000601
				US 2001-818875	A3 20010327
				WO 2001-US23770	W 20010727

AB Methods are presented for enhancing the efficiency of oligonucleotide-mediated repair or alteration of genetic information in cells having altered activity of DNA repair proteins using chimeric RNA-DNA double-stranded . The methods comprise using cells or cell-free exts. having altered levels or activity of at least one protein from the RAD52 epistasis group, the mismatch repair group or the nucleotide excision repair group. A assay system for identifying inhibitors of DNA repair proteins and monitoring genetic alteration using the oligonucleotides of the invention is also presented. Kits comprising cells and cell-free exts. having reduced activity of DNA repair proteins and vectors for enhancing targeted gene alteration are also presented. The invention demonstrates that gene repair depends on the dose of DNA repair proteins and expression of RAD52 gene suppresses oligonucleotide-directed gene alteration.

IT **394709-23-0**
RL: PRP (Properties)
(unclaimed sequence; methods for enhancing targeted gene alteration in cells having altered activity of DNA repair proteins using chimeric RNA-DNA double-stranded hairpin oligonucleotides)

L3 ANSWER 7 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 27 Jul 2001

ACCESSION NUMBER:	2001:545718 CAPLUS
DOCUMENT NUMBER:	135:149588
TITLE:	Method of affinity purifying proteins using modified bis-arsenical fluorescein
INVENTOR(S):	Vale, Ronald D.; Thorn, Kurt; Cooke, Roger; Matuska, Marija; Naber, Nariman
PATENT ASSIGNEE(S):	The Regents of the University of California, USA
SOURCE:	PCT Int. Appl., 52 pp. CODEN: PIXXD2
DOCUMENT TYPE:	Patent
LANGUAGE:	English
FAMILY ACC. NUM. COUNT:	1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE

WO 2001053325	A2	20010726	WO 2001-US2214		20010122
WO 2001053325	A3	20020307			
W: AU, CA, JP					
RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR					
US 6831160	B1	20041214	US 2000-502664		20000211
AU 2001031086	A5	20010731	AU 2001-31086		20010122
US 2005233428	A1	20051020	US 2004-12853		20041214
PRIORITY APPLN. INFO.:			US 2000-178054P	P	20000124
			US 2000-502664	A	20000211
			WO 2001-US2214	W	20010122

OTHER SOURCE(S): MARPAT 135:149588

AB The present invention features methods for purifying polypeptides of interest using a modified Fluorescein arsenical helix binder (FlAsH) compound immobilized on a solid support. An exemplary FlAsH target sequence motif is also presented. Examples of modification of the FlAsH compound which allow immobilization to a solid support are also provided. The present invention also provides DNA constructs for producing a dual affinity tagged polypeptide and methods for purification thereof. Human kinesin constructs C-terminally tagged with the peptide WEAAAREACCRECCARA (specifically chelating with β-alanine-modified FlAsH, preparation given) were expressed in Escherichia coli and purified using beads containing β-alanine-modified FlAsH. Protein was eluted using 1,2-ethanedithiol.

IT **268741-28-2P**
RL: BPN (Biosynthetic preparation); BPR (Biological process); BSU (Biological study, unclassified); NUU (Other use, unclassified); PRP (Properties); BIOL (Biological study); PREP (Preparation); PROC (Process); USES (Uses)
(amino acid sequence, as FlAsH peptide target; affinity purifying proteins using modified bis-arsenical fluorescein)

L3 ANSWER 8 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 02 Mar 2001
ACCESSION NUMBER: 2001:152863 CAPLUS
DOCUMENT NUMBER: 134:204756
TITLE: Methods for the detection, analysis and isolation of nascent proteins
INVENTOR(S): Rothschild, Kenneth J.; Gite, Sadanand; Olejnik, Jerzy
PATENT ASSIGNEE(S): Ambergen, Inc., USA
SOURCE: PCT Int. Appl., 204 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 3
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001014578	A1	20010301	WO 2000-US23233	20000823
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ,				

	UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
	RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
US 6303337		B1	20011016	US 1999-382950	19990825
US 6306628		B1	20011023	US 1999-382736	19990825
CA 2383554		AA	20010301	CA 2000-2383554	20000823
EP 1210449		A1	20020605	EP 2000-957758	20000823
	R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL				
JP 2004513604		T2	20040513	JP 2001-518889	20000823
AU 775940		B2	20040819	AU 2000-69328	20000823
US 2005009013		A1	20050113	US 2001-813197	20010320
US 6875592		B2	20050405		
US 2002132248		A1	20020919	US 2001-973145	20011009
US 2003092031		A1	20030515	US 2002-174368	20020618
US 2003190643		A1	20031009	US 2002-264127	20021003
US 2004014071		A1	20040122	US 2003-339712	20030110
US 2005032078		A1	20050210	US 2003-719523	20031121
PRIORITY APPLN. INFO.:				US 1999-382736	A 19990825
				US 1999-382950	A 19990825
				WO 2000-US23233	W 20000823
				US 2001-813197	A1 20010320
				US 2002-49332	A2 20020621

AB This invention relates to non-radioactive markers that facilitate the detection and anal. of nascent proteins translated within cellular or cell-free translation systems. Nascent proteins containing these markers can be rapidly and efficiently detected, isolated and analyzed without the handling and disposal problems associated with radioactive reagents. Preferred markers are dipyrrometheneboron difluoride (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene) dyes.

IT **268741-28-2**
RL: PRP (Properties)
(unclaimed sequence; methods for the detection, anal. and isolation of nascent proteins)

REFERENCE COUNT: 4 THERE ARE 4 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L3 ANSWER 9 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 18 Aug 2000

ACCESSION NUMBER:	2000:573678 CAPLUS
DOCUMENT NUMBER:	133:172215
TITLE:	Controlling protein levels in eucaryotic organisms using novel compds. comprising a ubiquitination recognition element and a protein binding element
INVENTOR(S):	Kenten, John H.; Roberts, Steven F.; Lebowitz, Michael S.
PATENT ASSIGNEE(S):	Proteinix, Inc., USA
SOURCE:	PCT Int. Appl., 106 pp. CODEN: PIXXD2
DOCUMENT TYPE:	Patent
LANGUAGE:	English

FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.		DATE
WO 2000047220	A1	20000817	WO 2000-US3436		20000211
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM					
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG					
US 6306663	B1	20011023	US 1999-406781		19990928
CA 2362560	AA	20000817	CA 2000-2362560		20000211
EP 1156817	A1	20011128	EP 2000-908580		20000211
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO					
JP 2002536417	T2	20021029	JP 2000-598172		20000211
US 2002146843	A1	20021010	US 2001-880149		20010614
US 2002173049	A1	20021121	US 2001-880132		20010614
US 6559280	B2	20030506			
US 2003153727	A1	20030814	US 2003-345281		20030116
PRIORITY APPLN. INFO.:			US 1999-119851P	P	19990212
			US 1999-406781	A2	19990928
			WO 2000-US3436	W	20000211
			US 2001-880132	A3	20010614

AB The invention relates to novel compds. comprising a ubiquitination recognition element and a protein binding element. The invention also relates to the use of said compds. for modulating the level and/or activity of a target protein. The compds. are useful for the treatment of diseases such as infections, inflammatory conditions, cancer and genetic diseases. The compds. are also useful as insecticides and herbicides.

IT **223673-79-8 268741-28-2**
RL: PRP (Properties)
(unclaimed sequence; controlling protein levels in eucaryotic organisms using novel compds. comprising a ubiquitination recognition element and a protein binding element)

REFERENCE COUNT: 9 THERE ARE 9 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L3 ANSWER 10 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 16 Mar 2000
ACCESSION NUMBER: 2000:169797 CAPLUS
DOCUMENT NUMBER: 132:344976
TITLE: A novel method of affinity-purifying proteins using a bis-arsenical fluorescein
AUTHOR(S): Thorn, Kurt S.; Naber, Nariman; Matuska, Marija; Vale, Ronald D.; Cooke, Roger
CORPORATE SOURCE: Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA,

94143, USA
SOURCE: Protein Science (2000), 9(2), 213-217
CODEN: PRCIEI; ISSN: 0961-8368
PUBLISHER: Cambridge University Press
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Genetically-encoded affinity tags constitute an important strategy for purifying proteins. Here, we have designed a novel affinity matrix based on the bis-arsenical fluorescein dye FlAsH, which specifically recognizes short α-helical peptides containing the sequence CCXXCC. We find that kinesin tagged with this cysteine-containing helix binds specifically to FlAsH resin and can be eluted in a fully active form. This affinity tag has several advantages over polyhistidine, the only small affinity tag in common use. The protein obtained with this single chromatog. step from crude Escherichia coli lysates is purer than that obtained with nickel affinity chromatog. of 6xHis tagged kinesin. Moreover, unlike nickel affinity chromatog., which requires high concns. of imidazole or pH changes for elution, protein bound to the FlAsH column can be completely eluted by dithiothreitol. Because of these mild elution conditions, FlAsH affinity chromatog. is ideal for recovering fully active protein and for the purification of intact protein complexes.
IT **268741-28-2**
RL: BPR (Biological process); BSU (Biological study, unclassified); BUU (Biological use, unclassified); BIOL (Biological study); PROC (Process); USES (Uses)
(novel method of affinity-purifying proteins using a bis-arsenical fluorescein)
REFERENCE COUNT: 10 THERE ARE 10 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L3 ANSWER 11 OF 11 CAPLUS COPYRIGHT 2005 ACS on STN
ED Entered STN: 10 May 1999
ACCESSION NUMBER: 1999:286159 CAPLUS
DOCUMENT NUMBER: 130:308804
TITLE: Target protein sequences for binding of synthetic biarsenical molecules
INVENTOR(S): Tsien, Roger Y.; Griffin, Albert B.
PATENT ASSIGNEE(S): The Regents of the University of California, USA
SOURCE: PCT Int. Appl., 77 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9921013	A1	19990429	WO 1998-US22363	19981021

W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, US, US, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

US 5932474	A	19990803	US 1997-955206	19971021
US 6008378	A	19991228	US 1997-955859	19971021
US 6054271	A	20000425	US 1997-955050	19971021
AU 9911139	A1	19990510	AU 1999-11139	19981021
EP 1032837	A1	20000906	EP 1998-953881	19981021
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
US 6451569	B1	20020917	US 1999-372338	19990811
US 2003083373	A1	20030501	US 2002-126752	20020419
US 6686458	B2	20040203		
US 2005131217	A1	20050616	US 2004-772164	20040203

PRIORITY APPLN. INFO.:

US 1997-955050	A2	19971021
US 1997-955206	A2	19971021
US 1997-955859	A2	19971021
WO 1998-US22363	W	19981021
US 1999-372338	A1	19990811
US 2002-126752	A1	20020419

OTHER SOURCE(S): MARPAT 130:308804

AB The present invention features biarsenical mols. and target sequences that specifically react with the biarsenical mols. A bonding partner comprises a carrier polypeptide and a target sequence, wherein the target sequence is heterologous to the carrier polypeptide and the target sequence contains one or more cysteines capable of specifically reacting with a biarsenical mol. Bonding partners that include target sequences, vectors that include nucleic acid sequences that encode the target sequences and host cells that include the target sequences are also featured in the invention. One example of a biarsenical compound is an arsenical derivative of fluorescein.

IT **223673-78-7**
RL: ARU (Analytical role, unclassified); BPR (Biological process); BSU (Biological study, unclassified); ANST (Analytical study); BIOL (Biological study); PROC (Process)
(SEQ ID 1; target protein sequences for binding of synthetic biarsenical mols.)

IT **223673-79-8**
RL: ARU (Analytical role, unclassified); BPR (Biological process); BSU (Biological study, unclassified); ANST (Analytical study); BIOL (Biological study); PROC (Process)
(SEQ ID 4; target protein sequences for binding of synthetic biarsenical mols.)

REFERENCE COUNT: 2 THERE ARE 2 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

FILE 'MEDLINE' ENTERED AT 12:36:52 ON 12 DEC 2005

FILE 'BIOSIS' ENTERED AT 12:36:52 ON 12 DEC 2005

FILE 'EMBASE' ENTERED AT 12:36:52 ON 12 DEC 2005

L4 0 L1

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 8, 2005, 15:48:28 ; Search time 229 Seconds
(without alignments)
52.375 Million cell updates/sec

Title: US-10-772-164-1
Perfect score: 101
Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	53	52.5	157	1	VE6_HPV12	P36803 human papil
2	51	50.5	278	2	Q6IGH0_DROME	Q6igh0 drosophila
3	51	50.5	502	2	Q9BGM9_9MAMM	Q9bgm9 tachyglossu
4	50	49.5	1370	1	ZN261_HUMAN	Q14202 homo sapien
5	50	49.5	1370	1	ZN261_MOUSE	Q9jlm4 mus musculu
6	49.5	49.0	602	2	Q75NZ5_CHLRE	Q75nz5 chlamydomon
7	49	48.5	115	1	ALK1_PIG	P22298 sus scrofa
8	49	48.5	155	2	Q9PXB1_HPV08	Q9pxb1 human papil
9	49	48.5	168	1	VE6_HPV21	P28832 human papil
10	49	48.5	1067	2	Q4QFE4_LEIMA	Q4qfe4 leishmania
11	48.5	48.0	390	2	Q4S604_TETNG	Q4s604 tetraodon n
12	48	47.5	62	2	Q4PN38_IXOSC	Q4pn38 ixodes scap
13	48	47.5	131	1	ALK1_MOUSE	P97430 mus musculu
14	48	47.5	131	2	Q548X8_MOUSE	Q548x8 mus musculu
15	48	47.5	157	2	O40617_HPVR7	O40617 human papil

16	48	47.5	157	2	Q81986_HPV05	Q81986 human papil
17	48	47.5	157	2	Q913V6_9PAPI	Q913v6 human papil
18	47.5	47.0	525	2	Q64FQ2_ARATH	Q64fq2 arabidopsis
19	47.5	47.0	676	2	O48785_ARATH	O48785 arabidopsis
20	47	46.5	88	2	Q62H93_BURMA	Q62h93 burkholderi
21	47	46.5	101	2	Q4IVT4_AZOVI	Q4ivt4 azotobacter
22	47	46.5	131	2	Q9R0Z8_RAT	Q9r0z8 rattus norv
23	47	46.5	156	1	VE6_HPV47	P22422 human papil
24	47	46.5	171	1	VE6_HPV14	P28830 human papil
25	47	46.5	181	2	Q8VMH1_PSEPU	Q8vmh1 pseudomonas
26	47	46.5	193	1	KR415_HUMAN	Q9byq5 homo sapien
27	47	46.5	210	1	KRA47_HUMAN	Q9byr0 homo sapien
28	47	46.5	219	2	Q52396_PSEST	Q52396 pseudomonas
29	47	46.5	1175	2	Q4P5X7_USTMA	Q4p5x7 ustilago ma
30	46	45.5	80	1	IBB4_LONCA	P16343 lonchocarpu
31	46	45.5	88	2	Q52509_PSESX	Q52509 pseudomonas
32	46	45.5	129	1	KRA56_HUMAN	Q618g9 homo sapien
33	46	45.5	161	2	Q8MZ55_DROME	Q8mz55 drosophila
34	46	45.5	166	1	VE6_HPV19	P36806 human papil
35	46	45.5	186	1	KRA45_HUMAN	Q9byr2 homo sapien
36	46	45.5	191	2	Q28583_SHEEP	Q28583 ovis aries
37	46	45.5	298	2	Q65T35_MANSM	Q65t35 mannheimia
38	46	45.5	412	2	P91666_DROME	P91666 drosophila
39	46	45.5	465	1	HYIN2_BRAJA	P19922 bradyrhizob
40	46	45.5	491	2	Q4T2B4_TETNG	Q4t2b4 tetraodon n
41	46	45.5	757	2	Q6PFS4_BRARE	Q6pfs4 brachydanio
42	46	45.5	1033	2	Q4T6W6_TETNG	Q4t6w6 tetraodon n
43	46	45.5	1063	2	Q4TBG6_TETNG	Q4tbg6 tetraodon n
44	46	45.5	1367	2	Q629H4_CAEBR	Q629h4 caenorhabdi
45	46	45.5	1376	2	Q23590_CAEEL	Q23590 caenorhabdi
46	46	45.5	1955	2	Q9VXG2_DROME	Q9vxg2 drosophila
47	46	45.5	1959	2	Q9VXG1_DROME	Q9vxg1 drosophila
48	45.5	45.0	139	2	Q8RYZ5_ORYSA	Q8ryz5 oryza sativ
49	45.5	45.0	279	2	Q4RZU3_TETNG	Q4rzu3 tetraodon n
50	45	44.6	61	2	Q9PB81_XYLFA	Q9pb81 xylella fas
51	45	44.6	68	2	O97751_PIG	O97751 sus scrofa
52	45	44.6	100	2	Q5UPC9_MIMIV	Q5upc9 mimivirus.
53	45	44.6	117	2	Q76YA2_9CAUD	Q76ya2 bacteriopha
54	45	44.6	120	2	Q9QQ85_HPV08	Q9qq85 human papil
55	45	44.6	130	2	Q6IE20_RAT	Q6ie20 rattus norv
56	45	44.6	149	2	O12671_9PAPI	O12671 colobus mon
57	45	44.6	155	1	VE6_HPV08	P06428 human papil
58	45	44.6	157	1	VE6_HPV05	P06930 human papil
59	45	44.6	157	1	VE6_HPV36	P50810 human papil
60	45	44.6	157	1	VE6_HPV5B	P26556 human papil
61	45	44.6	157	2	Q76WJ7_HPV5B	Q76wj7 human papil
62	45	44.6	157	2	Q6LBH6_HPV05	Q6lbh6 human papil
63	45	44.6	157	2	Q6YNY6_9PAPI	Q6yny6 human papil
64	45	44.6	157	2	Q81962_HPV05	Q81962 human papil
65	45	44.6	157	2	Q81985_HPV05	Q81985 human papil
66	45	44.6	157	2	Q910D7_9PAPI	Q910d7 human papil
67	45	44.6	157	2	Q910X3_9PAPI	Q910x3 human papil
68	45	44.6	157	2	Q913V7_9PAPI	Q913v7 human papil
69	45	44.6	157	2	Q913V8_9PAPI	Q913v8 human papil
70	45	44.6	157	2	Q913V9_9PAPI	Q913v9 human papil
71	45	44.6	157	2	Q913W0_9PAPI	Q913w0 human papil
72	45	44.6	157	2	Q913W1_9PAPI	Q913w1 human papil

73	45	44.6	157	2	Q913W2_9PAPI	Q913w2 human papil
74	45	44.6	157	2	Q913W3_9PAPI	Q913w3 human papil
75	45	44.6	157	2	Q913W4_9PAPI	Q913w4 human papil
76	45	44.6	157	2	Q913W5_9PAPI	Q913w5 human papil
77	45	44.6	157	2	Q913W6_9PAPI	Q913w6 human papil
78	45	44.6	165	1	VE6_HPV20	P28831 human papil
79	45	44.6	165	2	Q9D7P3_MOUSE	Q9d7p3 mus musculu
80	45	44.6	233	2	Q7RZM5_NEUCR	Q7rzm5 neurospora
81	45	44.6	369	2	Q9RVW7_DEIRA	Q9rvw7 deinococcus
82	45	44.6	485	2	Q698X7_9BRAS	Q698x7 thlaspi cae
83	45	44.6	485	2	Q8LST0_9BRAS	Q8lst0 thlaspi cae
84	45	44.6	485	2	Q8LST1_9BRAS	Q8lst1 thlaspi jap
85	45	44.6	533	2	Q4S3Z6_TETNG	Q4s3z6 tetraodon n
86	45	44.6	537	2	Q69TJ4_ORYSA	Q69tj4 oryza sativ
87	45	44.6	667	2	O77064_APLCA	O77064 aplysia cal
88	45	44.6	674	2	Q5DTK3_MOUSE	Q5dtk3 mus musculu
89	45	44.6	952	2	Q5KL94_CRYNE	Q5kl94 cryptococcu
90	45	44.6	1001	2	Q55UY8_CRYNE	Q55uy8 cryptococcu
91	44.5	44.1	134	2	Q6ZRE8_HUMAN	Q6zre8 homo sapien
92	44.5	44.1	411	2	Q5GUQ2_XANOR	Q5guq2 xanthomonas
93	44.5	44.1	416	2	Q522U4_MAGGR	Q522u4 magnaporthe
94	44.5	44.1	964	2	Q4STC1_TETNG	Q4stc1 tetraodon n
95	44	43.6	101	1	Y106_ENCCU	Q8st79 encephalito
96	44	43.6	101	1	Y702_ENCCU	Q8sv56 encephalito
97	44	43.6	161	1	VE6_HPV25	P28833 human papil
98	44	43.6	168	2	Q9D9I2_MOUSE	Q9d9i2 mus musculu
99	44	43.6	168	2	Q8CH20_MOUSE	Q8ch20 mus musculu
100	44	43.6	168	2	Q6P8T4_MOUSE	Q6p8t4 mus musculu

ALIGNMENTS

```
RESULT 1
VE6_HPV12
ID   VE6_HPV12      STANDARD;      PRT;   157 AA.
AC   P36803;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   E6 protein.
GN   Name=E6;
OS   Human papillomavirus type 12.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Betapapillomavirus.
OX   NCBI_TaxID=10604;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX   MEDLINE=94265501; PubMed=8205838;
RA   Delius H., Hofmann B.;
RT   "Primer-directed sequencing of human papillomavirus types.";
RL   Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC   -!- FUNCTION: Transcriptional transactivator. Binds double stranded
CC       DNA (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear matrix-associated (By similarity).
CC   -!- SIMILARITY: Belongs to the papillomaviruses E6 protein family.
CC   ---------------------------------------------------------------------
```

```
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   --------------------------------------------------------------------------
DR   EMBL; X74466; CAA52496.1; -; Genomic_DNA.
DR   PIR; S36538; S36538.
DR   InterPro; IPR001334; E6.
DR   Pfam; PF00518; E6; 1.
KW   Activator; DNA-binding; Early protein; Metal-binding; Nuclear protein;
KW   Transcription; Transcription regulation; Zinc; Zinc-finger.
FT   ZN_FING      39 .   75       Potential.
FT   ZN_FING     112    148       Potential.
SQ   SEQUENCE   157 AA;  17984 MW;  E9EC735537733FDC CRC64;

  Query Match             52.5%;  Score 53;  DB 1;  Length 157;
  Best Local Similarity   53.3%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 WEAAAREACCRECCA 15
              |:     |||| |||
Db         63 WKGHFVTACCRSCCA 77


RESULT 2
Q6IGH0_DROME
ID   Q6IGH0_DROME PRELIMINARY;      PRT;   278 AA.
AC   Q6IGH0;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   HDC06306.
GN   ORFNames=HDC06306;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=14709175; DOI=10.1186/gb-2003-5-1-r3;
RA   Hild M., Beckmann B., Haas S.A., Koch B., Solovyev V., Busold C.,
RA   Fellenberg K., Boutros M., Vingron M., Sauer F., Hoheisel J.D.,
RA   Paro R.;
RT   "An integrated gene annotation and transcriptional profiling approach
RT   towards the full gene content of the Drosophila genome.";
RL   Genome Biol. 5:RESEARCH0003.1-RESEARCH0003.17(2003).
CC   -!- MISCELLANEOUS: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ third party annotation (TPA) entry.
DR   EMBL; BK003796; DAA02494.1; -; Genomic_DNA.
DR   InterPro; IPR006209; EGF_like.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
SQ   SEQUENCE   278 AA;  32016 MW;  06E7253102FE5BF1 CRC64;

  Query Match             50.5%;  Score 51;  DB 2;  Length 278;
  Best Local Similarity   87.5%;  Pred. No. 35;
```

```
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          9 CCRECCAR 16
              |||||| |
Db        250 CCRECCCR 257



RESULT 8
Q9PXB1_HPV08
ID   Q9PXB1_HPV08  PRELIMINARY;      PRT;   155 AA.
AC   Q9PXB1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   E6 protein.
OS   Human papillomavirus type 8.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10579;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=91361540; PubMed=1653484;
RA   Deau M.C., Favre M., Orth G.;
RT   "Genetic heterogeneity among human papillomaviruses (HPV) associated
RT   with epidermodysplasia verruciformis: evidence for multiple allelic
RT   forms of HPV5 and HPV8 E6 genes.";
RL   Virology 184:492-503(1991).
DR   GO; GO:0042025; C:host cell nucleus; IEA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001334; E6.
DR   Pfam; PF00518; E6; 1.
SQ   SEQUENCE   155 AA;  17764 MW;  6986A0F88C7A33FD CRC64;

  Query Match             48.5%;  Score 49;  DB 2;  Length 155;
  Best Local Similarity   53.3%;  Pred. No. 41;
  Matches    8;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 WEAAAREACCRECCA 15
              |:     |||| |||
Db         63 WKNYVVTACCRCCCA 77



RESULT 30
IBB4_LONCA
ID   IBB4_LONCA     STANDARD;      PRT;    80 AA.
AC   P16343;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Bowman-Birk type proteinase inhibitor DE-4 (DE4).
OS   Lonchocarpus capassa (Apple-leaf).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
```

```
OC   rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Millettieae;
OC   Lonchocarpus.
OX   NCBI_TaxID=3926;
RN   [1]
RP   PROTEIN SEQUENCE.
RC   TISSUE=Seed;
RA   Joubert F.J.;
RT   "Proteinase inhibitors from Lonchocarpus capassa (apple-leaf) seed.";
RL   Phytochemistry 23:957-961(1984).
CC   -!- SIMILARITY: Belongs to the Bowman-Birk serine protease inhibitor
CC       family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   HSSP; P01062; 1DF9.
DR   InterPro; IPR000877; Prot_inh_BBI.
DR   Pfam; PF00228; Bowman-Birk_leg; 2.
DR   SMART; SM00269; BowB; 1.
DR   PROSITE; PS00281; BOWMAN_BIRK; 1.
KW   Direct protein sequencing; Protease inhibitor;
KW   Serine protease inhibitor.
FT   SITE         25     26       Reactive bond for trypsin (By
FT                                similarity).
FT   SITE         52     53       Reactive bond for chymotrypsin (By
FT                                similarity).
FT   DISULFID     18     71       By similarity.
FT   DISULFID     19     33       By similarity.
FT   DISULFID     22     67       By similarity.
FT   DISULFID     23     31       By similarity.
FT   DISULFID     41     48       By similarity.
FT   DISULFID     45     60       By similarity.
FT   DISULFID     50     58       By similarity.
SQ   SEQUENCE   80 AA;  8806 MW;  6E8DF76866B871C9 CRC64;

  Query Match             45.5%;  Score 46;  DB 1;  Length 80;
  Best Local Similarity   37.5%;  Pred. No. 60;
  Matches    6;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 EAAAREACCRECCARA 17
              |: : : ||  || |:
Db         11 ESESSKPCCSSCCTRS 26


Search completed: December  8, 2005, 16:07:48
Job time : 233 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 8, 2005, 15:48:56 ; Search time 37 Seconds
(without alignments)
44.208 Million cell updates/sec

Title: US-10-772-164-1
Perfect score: 101
Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	53	52.5	157	2	S36538	E6 protein - human
2	49	48.5	115	2	A36113	antileukoproteinas
3	47.5	47.0	676	2	G84663	hypothetical prote
4	47	46.5	156	1	W6WL47	E6 protein - human
5	46	45.5	166	2	S36485	E6 protein - human
6	46	45.5	191	2	I46412	keratin KAP5.4 - s
7	46	45.5	465	2	S05311	indoleacetamide hy
8	46	45.5	498	2	A48203	interleukin-14 pre
9	46	45.5	571	2	S69210	protein kinase cak
10	46	45.5	1430	2	T34516	hypothetical prote
11	45	44.6	61	2	E82580	hypothetical prote
12	45	44.6	155	1	W6WL8	E6 protein - human
13	45	44.6	157	1	W6WL5	E6 protein - human

14	45	44.6	157	1	W6WLB5	E6 protein - human
15	45	44.6	273	2	A43862	29K peripheral mem
16	45	44.6	369	2	G75460	hypothetical prote
17	44	43.6	161	2	S36491	E6 protein - human
18	44	43.6	186	2	A45910	ultra-high-sulfur
19	44	43.6	188	2	JC6547	high sulfur protei
20	44	43.6	204	2	T08072	proteinase inhibit
21	44	43.6	251	2	AH3413	nitrogen fixation
22	44	43.6	254	2	B84901	hypothetical prote
23	44	43.6	299	2	C97102	hypothetical prote
24	44	43.6	370	1	S57347	Ca2+/calmodulin-de
25	44	43.6	374	1	S50193	Ca2+/calmodulin-de
26	44	43.6	496	2	F75257	hypothetical prote
27	44	43.6	994	2	A48849	Ca2+-transporting
28	44	43.6	1001	1	PWRBFC	Ca2+-transporting
29	44	43.6	1121	2	S30862	DNA dependent ATPa
30	43.5	43.1	126	2	I46489	cysteine-rich hair
31	43	42.6	169	1	S18946	ultra high-sulfur
32	43	42.6	217	2	T33353	hypothetical prote
33	43	42.6	221	2	C34768	ORF2 protein - Orf
34	43	42.6	233	2	S67947	alkyl hydroperoxid
35	43	42.6	399	2	B24698	formate dehydrogen
36	43	42.6	689	2	T08988	cadmium-transporti
37	43	42.6	711	2	A85352	cadmium-transporti
38	43	42.6	976	2	D96714	DNA-directed RNA p
39	42.5	42.1	931	2	H96527	protein F27J15.16
40	42	41.6	122	2	JC6548	high sulfur protei
41	42	41.6	223	2	B38346	ultra-high-sulfur
42	42	41.6	230	2	A38346	ultra-high-sulfur
43	42	41.6	247	2	T17311	hypothetical prote
44	42	41.6	327	2	C86452	protein F6N18.11 [
45	42	41.6	1212	2	B82809	exodeoxyribonuclea
46	42	41.6	2037	2	T16881	hypothetical prote
47	41	40.6	67	2	T37199	hypothetical prote
48	41	40.6	151	2	S60314	hair keratin cyste
49	41	40.6	164	2	T24272	hypothetical prote
50	41	40.6	169	2	T06062	hypothetical prote
51	41	40.6	188	2	T15651	hypothetical prote
52	41	40.6	199	2	T48099	hypothetical prote
53	41	40.6	211	2	H71281	probable endonucle
54	41	40.6	215	2	G86255	protein F12F1.7 [i
55	41	40.6	352	2	S11926	cellulose 1,4-beta
56	41	40.6	369	2	F69407	iron-sulfur cluste
57	41	40.6	452	2	G86170	hypothetical prote
58	41	40.6	508	2	T22836	hypothetical prote
59	41	40.6	907	2	T02417	probable C2H2-type
60	41	40.6	997	2	S33754	glutamate receptor
61	40.5	40.1	229	2	S60454	glucose starvation
62	40	39.6	51	2	S78712	protein YDR034w-b
63	40	39.6	63	2	S00951	hypothetical prote
64	40	39.6	113	2	T03966	allergenic protein
65	40	39.6	130	2	F72513	hypothetical prote
66	40	39.6	132	1	TIHUSP	antileukoproteinas
67	40	39.6	152	2	T18975	hypothetical prote
68	40	39.6	174	2	S71554	pathogenesis-relat
69	40	39.6	181	2	A86451	probable ferredoxi
70	40	39.6	264	2	JC6125	U2 small nuclear r

71	40	39.6	548	2	C86456	unknown protein [i
72	40	39.6	619	2	C96714	unknown protein T6
73	40	39.6	708	2	T00064	hypothetical prote
74	40	39.6	709	2	T28712	hypothetical prote
75	40	39.6	860	2	A96717	unknown protein, 4
76	40	39.6	898	2	A69092	alanine-tRNA ligas
77	40	39.6	898	2	A40114	fasciclin II precu
78	40	39.6	1112	2	S28289	hypothetical prote
79	40	39.6	1385	2	A88554	protein C38C10.5a
80	40	39.6	1391	2	B88554	protein C38C10.5b
81	40	39.6	2523	2	T18477	hypothetical prote
82	39.5	39.1	26	2	C39414	electron transport
83	39.5	39.1	138	2	T25620	hypothetical prote
84	39.5	39.1	300	2	T03464	probable methylene
85	39.5	39.1	498	2	B69276	hypothetical prote
86	39.5	39.1	893	2	T38147	dolichyl-phosphate
87	39	38.6	55	2	E70593	probable rubA prot
88	39	38.6	81	1	TIZB2	proteinase inhibit
89	39	38.6	136	2	S78428	destabilase 2 - me
90	39	38.6	171	2	S35248	nifQ protein - Ent
91	39	38.6	173	1	RUPSEO	rubredoxin II - Ps
92	39	38.6	200	2	JC6068	U2 auxiliary facto
93	39	38.6	215	2	T39341	hypothetical prote
94	39	38.6	216	2	T39243	splicing factor u2
95	39	38.6	240	2	A46179	U2 snRNP auxiliary
96	39	38.6	287	2	A41257	apoptosis protein
97	39	38.6	332	2	JC1229	adenosine receptor
98	39	38.6	343	2	I49067	zinc finger protei
99	39	38.6	344	2	I52969	programmed cell de
100	39	38.6	352	2	T47820	hypothetical prote

ALIGNMENTS

RESULT 1
S36538
E6 protein - human papillomavirus type 12
C;Species: human papillomavirus type 12
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C;Accession: S36538
R;Delius, H.; Hofmann, B.
submitted to the EMBL Data Library, August 1993
A;Description: Primer-directed sequencing of human papillomavirus types.
A;Reference number: S36469
A;Accession: S36538
A;Molecule type: DNA
A;Residues: 1-157 <DEL>
A;Cross-references: UNIPROT:P36803; UNIPARC:UPI00001383B8; EMBL:X74466;
NID:g396910; PIDN:CAA52496.1; PID:g396911
C;Superfamily: papillomavirus E6 protein
C;Keywords: DNA binding; early protein; nucleus; zinc finger

Query Match 52.5%; Score 53; DB 2; Length 157;
Best Local Similarity 53.3%; Pred. No. 2.9;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```
Qy          1 WEAAAREACCRECCA 15
              |:     |||| |||
Db         63 WKGHFVTACCRSCCA 77
```

RESULT 2
A36113
antileukoproteinase precursor - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 28-Mar-1991 #sequence_revision 13-Jan-1993 #text_change 09-Jul-2004
C;Accession: A36113; A49198
R;Farmer, S.J.; Fliss, A.E.; Simmen, R.C.M.
Mol. Endocrinol. 4, 1095-1104, 1990
A;Title: Complementary DNA cloning and regulation of expression of the messenger RNA encoding a pregnancy-associated porcine uterine protein related to human antileukoproteinase.
A;Reference number: A36113; MUID:91155942; PMID:2293019
A;Accession: A36113
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-115 <FAR>
A;Cross-references: UNIPROT:P22298; UNIPARC:UPI0000125858; GB:M57446; NID:g164319; PIDN:AAA63446.1; PID:g164320
A;Note: the authors translated the codon GCT for residue 52 as Gly
R;Simmen, R.C.; Michel, F.J.; Fliss, A.E.; Smith, L.C.; Fliss, M.F.
Endocrinology 130, 1957-1965, 1992
A;Title: Ontogeny, immunocytochemical localization, and biochemical properties of the pregnancy-associated uterine elastase/cathepsin-G protease inhibitor, antileukoproteinase (ALP): monospecific antibodies to a synthetic peptide recognize native ALP.
A;Reference number: A49198; MUID:92191891; PMID:1547723
A;Accession: A49198
A;Status: preliminary
A;Molecule type: protein
A;Residues: 9-26 <SIM>
A;Cross-references: UNIPARC:UPI0000087C99
A;Experimental source: uterus
A;Note: sequence extracted from NCBI backbone (NCBIP:89471)
C;Superfamily: antileukoproteinase; antileukoproteinase repeat homology
F;14-59/Domain: antileukoproteinase repeat homology <ALP1>
F;68-113/Domain: antileukoproteinase repeat homology <ALP2>

  Query Match   48.5%; Score 49; DB 2; Length 115;
  Best Local Similarity   40.0%; Pred. No. 8;
  Matches   6; Conservative   4; Mismatches   5; Indels   0; Gaps   0;

```
Qy          1 WEAAAREACCRECCA 15
              |:    :: |||: ||
Db         38 WQCPDKKKCCRDTCA 52
```

RESULT 4
W6WL47
E6 protein - human papillomavirus type 47
C;Species: human papillomavirus type 47
A;Note: host Homo sapiens (man)

C;Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 09-Jul-2004
C;Accession: A35324
R;Kiyono, T.; Adachi, A.; Ishibashi, M.
Virology 177, 401-405, 1990
A;Title: Genome organization and taxonomic position of human papillomavirus type 47 inferred from its DNA sequence.
A;Reference number: A35324; MUID:90281611; PMID:2162112
A;Accession: A35324
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-156 <KIY>
A;Cross-references: UNIPROT:P22422; UNIPARC:UPI00001383D9; GB:M32305; NID:g333062; PIDN:AAA46976.1; PID:g333064
C;Superfamily: papillomavirus E6 protein
C;Keywords: DNA binding; early protein; transforming protein; zinc finger
F;40-76/Region: zinc finger CCCC motif
F;113-149/Region: zinc finger CCCC motif

  Query Match   46.5%; Score 47; DB 1; Length 156;
  Best Local Similarity   46.7%; Pred. No. 18;
  Matches   7; Conservative   3; Mismatches   5; Indels   0; Gaps   0;

Qy   1 WEAAAREACCRECCA 15
              |: : |||| ||:
Db   64 WKDYSVYACCRLCCS 78

RESULT 6
I46412
keratin KAP5.4 - sheep
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
C;Accession: I46412; S34215
R;Jenkins, B.J.; Powell, B.C.
J. Invest. Dermatol. 103, 310-317, 1994
A;Title: Differential expression of genes encoding a cysteine-rich keratin family in the hair cuticle.
A;Reference number: I46412; MUID:94358466; PMID:7521375
A;Accession: I46412
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-191 <JEN>
A;Cross-references: UNIPROT:Q28583; UNIPARC:UPI0000088B92; EMBL:X73434; NID:g313719; PIDN:CAA51829.1; PID:g313720
C;Genetics:
A;Gene: KRTAP5.4
C;Superfamily: ultra-high-sulfur keratin

  Query Match   45.5%; Score 46; DB 2; Length 191;
  Best Local Similarity   38.5%; Pred. No. 28;
  Matches   5; Conservative   6; Mismatches   2; Indels   0; Gaps   0;

Qy   5 AREACCRECCARA 17
              :: :||| ||:::
Db   167 SQSSCCRPCCSQS 179
Search completed: December 8, 2005, 16:08:31
Job time : 40 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 8, 2005, 16:07:58 ; Search time 12 Seconds
(without alignments)
7.911 Million cell updates/sec

Title: US-10-772-164-1
Perfect score: 101
Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 32527 seqs, 5584426 residues

Total number of hits satisfying chosen parameters: 32527

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	44.6	1129	7	US-11-077-550-42	Sequence 42, Appl
2	45	44.6	1129	7	US-11-077-550-48	Sequence 48, Appl
3	45	44.6	1129	7	US-11-077-550-52	Sequence 52, Appl
4	45	44.6	1129	7	US-11-077-550-56	Sequence 56, Appl
5	45	44.6	1132	7	US-11-077-550-46	Sequence 46, Appl
6	43	42.6	3500	7	US-11-085-775-2	Sequence 2, Appli
7	42	41.6	75	6	US-10-478-345-12	Sequence 12, Appl
8	42	41.6	357	6	US-10-478-345-6	Sequence 6, Appli
9	41	40.6	720	7	US-11-102-240-38	Sequence 38, Appl

10	40	39.6	321	6	US-10-478-345-8	Sequence 8, Appli
11	39.5	39.1	898	7	US-11-174-150-43	Sequence 43, Appl
12	39	38.6	211	6	US-10-821-234-1372	Sequence 1372, Ap
13	39	38.6	998	6	US-10-510-524-1	Sequence 1, Appli
14	38	37.6	544	6	US-10-980-388-40	Sequence 40, Appl
15	38	37.6	831	6	US-10-467-657-4486	Sequence 4486, Ap
16	38	37.6	1907	7	US-11-039-398-25	Sequence 25, Appl
17	37.5	37.1	247	6	US-10-632-150-36	Sequence 36, Appl
18	37.5	37.1	247	7	US-11-073-457-36	Sequence 36, Appl
19	37.5	37.1	511	7	US-11-012-762-66	Sequence 66, Appl
20	37	36.6	133	7	US-11-047-757-9	Sequence 9, Appli
21	37	36.6	145	6	US-10-467-657-4246	Sequence 4246, Ap
22	37	36.6	179	6	US-10-467-657-2446	Sequence 2446, Ap
23	37	36.6	185	7	US-11-147-047-41	Sequence 41, Appl
24	37	36.6	376	7	US-11-116-939-8	Sequence 8, Appli
25	37	36.6	575	6	US-10-980-388-46	Sequence 46, Appl
26	36.5	36.1	1036	6	US-10-131-826A-142	Sequence 142, App
27	36	35.6	336	6	US-10-478-345-4	Sequence 4, Appli
28	36	35.6	393	6	US-10-821-234-1292	Sequence 1292, Ap
29	36	35.6	548	7	US-11-137-465-47	Sequence 47, Appl
30	36	35.6	1211	7	US-11-186-284-4	Sequence 4, Appli
31	35.5	35.1	277	7	US-11-132-285-3	Sequence 3, Appli
32	35.5	35.1	277	7	US-11-182-946-12	Sequence 12, Appl
33	35.5	35.1	1187	6	US-10-821-234-955	Sequence 955, App
34	35	34.7	211	6	US-10-467-657-8552	Sequence 8552, Ap
35	35	34.7	263	6	US-10-131-826A-484	Sequence 484, App
36	35	34.7	263	6	US-10-821-234-1403	Sequence 1403, Ap
37	35	34.7	346	6	US-10-878-556A-121	Sequence 121, App
38	35	34.7	346	7	US-11-069-642-109	Sequence 109, App
39	35	34.7	748	6	US-10-821-234-1479	Sequence 1479, Ap
40	35	34.7	750	7	US-11-089-551A-32	Sequence 32, Appl
41	34.5	34.2	228	6	US-10-980-388-17	Sequence 17, Appl
42	34.5	34.2	357	7	US-11-108-528-60	Sequence 60, Appl
43	34	33.7	177	6	US-10-821-234-1466	Sequence 1466, Ap
44	34	33.7	250	6	US-10-131-826A-320	Sequence 320, App
45	34	33.7	285	6	US-10-131-826A-448	Sequence 448, App
46	34	33.7	321	7	US-11-102-240-10	Sequence 10, Appl
47	34	33.7	337	6	US-10-467-962B-97	Sequence 97, Appl
48	34	33.7	440	7	US-11-102-240-134	Sequence 134, App
49	34	33.7	816	7	US-11-090-439-48	Sequence 48, Appl
50	34	33.7	821	7	US-11-087-227-90	Sequence 90, Appl
51	34	33.7	1076	6	US-10-131-826A-219	Sequence 219, App
52	33.5	33.2	75	6	US-10-478-345-14	Sequence 14, Appl
53	33.5	33.2	362	7	US-11-012-762-30	Sequence 30, Appl
54	33.5	33.2	362	7	US-11-012-762-32	Sequence 32, Appl
55	33.5	33.2	820	6	US-10-858-730-211	Sequence 211, App
56	33.5	33.2	1028	7	US-11-067-121-7	Sequence 7, Appli
57	33	32.7	23	6	US-10-967-457-73	Sequence 73, Appl
58	33	32.7	38	7	US-11-119-683-5	Sequence 5, Appli
59	33	32.7	46	6	US-10-467-657-7732	Sequence 7732, Ap
60	33	32.7	75	6	US-10-467-657-8898	Sequence 8898, Ap
61	33	32.7	106	7	US-11-020-772-20	Sequence 20, Appl
62	33	32.7	126	7	US-11-113-424-184	Sequence 184, App
63	33	32.7	148	6	US-10-526-716-2	Sequence 2, Appli
64	33	32.7	185	6	US-10-821-234-1130	Sequence 1130, Ap
65	33	32.7	231	6	US-10-467-657-2996	Sequence 2996, Ap
66	33	32.7	337	6	US-10-467-657-4674	Sequence 4674, Ap

67	33	32.7	354	6	US-10-478-345-2	Sequence 2, Appli
68	33	32.7	459	6	US-10-821-234-896	Sequence 896, App
69	33	32.7	567	6	US-10-467-657-4328	Sequence 4328, Ap
70	33	32.7	640	6	US-10-131-826A-368	Sequence 368, App
71	33	32.7	776	6	US-10-925-970-3	Sequence 3, Appli
72	33	32.7	928	6	US-10-841-129-4	Sequence 4, Appli
73	33	32.7	928	6	US-10-841-129-6	Sequence 6, Appli
74	33	32.7	1126	7	US-11-075-185-3	Sequence 3, Appli
75	33	32.7	1255	6	US-10-770-726-62	Sequence 62, Appl
76	33	32.7	1255	7	US-11-022-562-213	Sequence 213, App
77	32.5	32.2	148	6	US-10-526-716-4	Sequence 4, Appli
78	32.5	32.2	196	6	US-10-821-234-1682	Sequence 1682, Ap
79	32.5	32.2	480	6	US-10-821-234-1465	Sequence 1465, Ap
80	32.5	32.2	540	6	US-10-821-234-1456	Sequence 1456, Ap
81	32	31.7	10	7	US-11-053-076-228	Sequence 228, App
82	32	31.7	27	7	US-11-052-168A-20	Sequence 20, Appl
83	32	31.7	27	7	US-11-052-168A-25	Sequence 25, Appl
84	32	31.7	72	7	US-11-176-868-15	Sequence 15, Appl
85	32	31.7	95	7	US-11-119-212-11	Sequence 11, Appl
86	32	31.7	95	7	US-11-119-212-23	Sequence 23, Appl
87	32	31.7	98	7	US-11-082-381-11	Sequence 11, Appl
88	32	31.7	101	7	US-11-082-381-1	Sequence 1, Appli
89	32	31.7	185	7	US-11-179-411-20	Sequence 20, Appl
90	32	31.7	269	6	US-10-972-587-16	Sequence 16, Appl
91	32	31.7	270	6	US-10-972-587-14	Sequence 14, Appl
92	32	31.7	302	6	US-10-878-556A-61	Sequence 61, Appl
93	32	31.7	302	7	US-11-119-212-13	Sequence 13, Appl
94	32	31.7	302	7	US-11-119-212-25	Sequence 25, Appl
95	32	31.7	328	6	US-10-821-234-1671	Sequence 1671, Ap
96	32	31.7	344	6	US-10-967-527A-24	Sequence 24, Appl
97	32	31.7	368	7	US-11-178-737-3	Sequence 3, Appli
98	32	31.7	382	7	US-11-179-411-22	Sequence 22, Appl
99	32	31.7	411	7	US-11-119-212-17	Sequence 17, Appl
100	32	31.7	413	7	US-11-119-212-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-11-077-550-42
; Sequence 42, Application US/11077550
; Publication No. US20050244435A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130004
; CURRENT APPLICATION NUMBER: US/11/077,550
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: 10/241,596

; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
;   LENGTH: 1129
;   TYPE: PRT
;   ORGANISM: Clostridium botulinum
US-11-077-550-42

  Query Match   44.6%; Score 45; DB 7; Length 1129;
  Best Local Similarity   56.2%; Pred. No. 9.7;
  Matches   9; Conservative   3; Mismatches   4; Indels   0; Gaps   0;

Qy   2 EAAAREACCRECCARA 17
              ||||:|| :| |:|
Db   873 EAAAKEAAAKEAAAKA 888

RESULT 6
US-11-085-775-2
; Sequence 2, Application US/11085775
; Publication No. US20050260634A1
; GENERAL INFORMATION:
; APPLICANT: BALDWIN, DARYL
; APPLICANT: CLARK, HILARY
; APPLICANT: JUBB, ADRIAN
; APPLICANT: KOEPPEN, HARTMUT
; APPLICANT: QUAN, CLIFFORD
; APPLICANT: WU, THOMAS
; APPLICANT: ZHANG, ZEMIN
; TITLE OF INVENTION: ACHAETE-SCUTE LIKE-2 POLYPEPTIDES AND ENCODING NUCLEIC
; TITLE OF INVENTION: ACIDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF
TUMOR
; FILE REFERENCE: P5028R1P1-US
; CURRENT APPLICATION NUMBER: US/11/085,775
; CURRENT FILING DATE: 2005-03-21
; PRIOR APPLICATION NUMBER: PCT/US03/17682
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: US 10/454,945
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: US 60/407,087
; PRIOR FILING DATE: 2002-08-29
; NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 2
;   LENGTH: 3500
;   TYPE: PRT
;   ORGANISM: Homo sapiens

US-11-085-775-2

Query Match 42.6%; Score 43; DB 7; Length 3500;
Best Local Similarity 63.6%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          4 AAREACCRECC 14
              ||  |||  ||
Db       1274 AAAAACCAGCC 1284
```

Search completed: December 8, 2005, 16:19:27
Job time : 13 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 8, 2005, 16:08:38 ; Search time 162 Seconds
(without alignments)
43.846 Million cell updates/sec

Title: US-10-772-164-1
Perfect score: 101
Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	17	3	US-09-973-145-3	Sequence 3, Appli
2	101	100.0	17	3	US-09-880-149-48	Sequence 48, Appl
3	101	100.0	17	3	US-09-880-132-48	Sequence 48, Appl
4	101	100.0	17	3	US-09-813-197-4	Sequence 4, Appli
5	101	100.0	17	4	US-10-126-752-1	Sequence 1, Appli
6	101	100.0	17	4	US-10-174-368A-3	Sequence 3, Appli
7	101	100.0	17	4	US-10-345-281-48	Sequence 48, Appl
8	101	100.0	17	4	US-10-264-127-4	Sequence 4, Appli
9	101	100.0	17	4	US-10-339-712-4	Sequence 4, Appli
10	101	100.0	17	5	US-10-719-523-4	Sequence 4, Appli
11	101	100.0	17	5	US-10-772-164-1	Sequence 1, Appli

12	101	100.0	17	5	US-10-957-433-8	Sequence 8, Appli
13	101	100.0	17	6	US-11-012-853-2	Sequence 2, Appli
14	90	89.1	17	3	US-09-880-149-49	Sequence 49, Appl
15	90	89.1	17	3	US-09-880-132-49	Sequence 49, Appl
16	90	89.1	17	4	US-10-126-752-4	Sequence 4, Appli
17	90	89.1	17	4	US-10-345-281-49	Sequence 49, Appl
18	90	89.1	17	5	US-10-772-164-4	Sequence 4, Appli
19	87	86.1	19	3	US-09-818-875-4368	Sequence 4368, Ap
20	87	86.1	19	4	US-10-260-375A-16	Sequence 16, Appl
21	87	86.1	19	4	US-10-351-662-16	Sequence 16, Appl
22	87	86.1	19	4	US-10-209-787-4368	Sequence 4368, Ap
23	87	86.1	19	4	US-10-307-005-2700	Sequence 2700, Ap
24	87	86.1	19	4	US-10-261-185-4368	Sequence 4368, Ap
25	81	80.2	19	4	US-10-384-918-16	Sequence 16, Appl
26	54	53.5	4277	4	US-10-184-644-439	Sequence 439, App
27	54	53.5	4277	4	US-10-184-634-439	Sequence 439, App
28	53	52.5	189	4	US-10-437-963-149015	Sequence 149015,
29	53	52.5	2974	4	US-10-184-644-521	Sequence 521, App
30	53	52.5	2974	4	US-10-184-634-521	Sequence 521, App
31	50	49.5	28	4	US-10-252-136-14	Sequence 14, Appl
32	50	49.5	28	4	US-10-351-641-231	Sequence 231, App
33	50	49.5	28	4	US-10-267-682-161	Sequence 161, App
34	50	49.5	28	4	US-10-267-748-161	Sequence 161, App
35	50	49.5	152	4	US-10-767-701-60750	Sequence 60750, A
36	50	49.5	284	4	US-10-437-963-199693	Sequence 199693,
37	50	49.5	823	4	US-10-123-155-379	Sequence 379, App
38	50	49.5	823	4	US-10-146-731-379	Sequence 379, App
39	50	49.5	823	4	US-10-140-472-379	Sequence 379, App
40	50	49.5	823	4	US-10-141-761-379	Sequence 379, App
41	50	49.5	823	4	US-10-142-885-379	Sequence 379, App
42	50	49.5	823	4	US-10-158-790-379	Sequence 379, App
43	50	49.5	823	4	US-10-137-871-379	Sequence 379, App
44	50	49.5	823	4	US-10-140-923-379	Sequence 379, App
45	50	49.5	823	4	US-10-141-756-379	Sequence 379, App
46	50	49.5	823	4	US-10-141-759-379	Sequence 379, App
47	50	49.5	823	4	US-10-140-805-379	Sequence 379, App
48	50	49.5	823	4	US-10-140-864-379	Sequence 379, App
49	50	49.5	2012	4	US-10-437-963-204172	Sequence 204172,
50	50	49.5	2826	4	US-10-334-703-50	Sequence 50, Appl
51	49	48.5	285	4	US-10-017-161-1530	Sequence 1530, Ap
52	49	48.5	1346	4	US-10-123-155-481	Sequence 481, App
53	49	48.5	1346	4	US-10-146-731-481	Sequence 481, App
54	49	48.5	1346	4	US-10-140-472-481	Sequence 481, App
55	49	48.5	1346	4	US-10-141-761-481	Sequence 481, App
56	49	48.5	1346	4	US-10-142-885-481	Sequence 481, App
57	49	48.5	1346	4	US-10-158-790-481	Sequence 481, App
58	49	48.5	1346	4	US-10-137-871-481	Sequence 481, App
59	49	48.5	1346	4	US-10-140-923-481	Sequence 481, App
60	49	48.5	1346	4	US-10-141-756-481	Sequence 481, App
61	49	48.5	1346	4	US-10-141-759-481	Sequence 481, App
62	49	48.5	1346	4	US-10-140-805-481	Sequence 481, App
63	49	48.5	1346	4	US-10-140-864-481	Sequence 481, App
64	49	48.5	2076	4	US-10-184-644-409	Sequence 409, App
65	49	48.5	2076	4	US-10-184-634-409	Sequence 409, App
66	49	48.5	2212	4	US-10-184-644-325	Sequence 325, App
67	49	48.5	2212	4	US-10-184-634-325	Sequence 325, App
68	49	48.5	2623	4	US-10-123-155-451	Sequence 451, App

69	49	48.5	2623	4	US-10-146-731-451	Sequence 451, App
70	49	48.5	2623	4	US-10-140-472-451	Sequence 451, App
71	49	48.5	2623	4	US-10-141-761-451	Sequence 451, App
72	49	48.5	2623	4	US-10-142-885-451	Sequence 451, App
73	49	48.5	2623	4	US-10-158-790-451	Sequence 451, App
74	49	48.5	2623	4	US-10-137-871-451	Sequence 451, App
75	49	48.5	2623	4	US-10-140-923-451	Sequence 451, App
76	49	48.5	2623	4	US-10-141-756-451	Sequence 451, App
77	49	48.5	2623	4	US-10-141-759-451	Sequence 451, App
78	49	48.5	2623	4	US-10-140-805-451	Sequence 451, App
79	49	48.5	2623	4	US-10-140-864-451	Sequence 451, App
80	49	48.5	2782	4	US-10-123-155-205	Sequence 205, App
81	49	48.5	2782	4	US-10-146-731-205	Sequence 205, App
82	49	48.5	2782	4	US-10-140-472-205	Sequence 205, App
83	49	48.5	2782	4	US-10-141-761-205	Sequence 205, App
84	49	48.5	2782	4	US-10-142-885-205	Sequence 205, App
85	49	48.5	2782	4	US-10-158-790-205	Sequence 205, App
86	49	48.5	2782	4	US-10-137-871-205	Sequence 205, App
87	49	48.5	2782	4	US-10-140-923-205	Sequence 205, App
88	49	48.5	2782	4	US-10-141-756-205	Sequence 205, App
89	49	48.5	2782	4	US-10-141-759-205	Sequence 205, App
90	49	48.5	2782	4	US-10-140-805-205	Sequence 205, App
91	49	48.5	2782	4	US-10-140-864-205	Sequence 205, App
92	49	48.5	2846	4	US-10-184-644-169	Sequence 169, App
93	49	48.5	2846	4	US-10-184-634-169	Sequence 169, App
94	49	48.5	2846	4	US-10-063-685-37	Sequence 37, Appl
95	49	48.5	3004	4	US-10-184-644-91	Sequence 91, Appl
96	49	48.5	3004	4	US-10-184-634-91	Sequence 91, Appl
97	48	47.5	131	3	US-09-790-264-61	Sequence 61, Appl
98	48	47.5	131	4	US-10-269-353-61	Sequence 61, Appl
99	48	47.5	131	4	US-10-250-959-3	Sequence 3, Appli
100	48	47.5	131	5	US-10-900-926-61	Sequence 61, Appl

ALIGNMENTS

```
RESULT 1
US-09-973-145-3
; Sequence 3, Application US/09973145
; Patent No. US20020132248A1
; GENERAL INFORMATION:
;  APPLICANT: Rothschild, Kenneth J.
;  APPLICANT:  Gite, Sadanand
;  APPLICANT:  Olejnik, Jerzy
;  TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
;  FILE REFERENCE: AMBER-06819
;  CURRENT APPLICATION NUMBER: US/09/973,145
;  CURRENT FILING DATE:  2001-10-09
;  PRIOR APPLICATION NUMBER: 09/382,950
;  PRIOR FILING DATE: 1999-08-25
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 17
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
```

; FEATURE:
; OTHER INFORMATION: Synthetic
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Synthetic
US-09-973-145-3

  Query Match   100.0%; Score 101; DB 3; Length 17;
  Best Local Similarity   100.0%; Pred. No. 1.2e-05;
  Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 WEAAAREACCRECCARA 17
              |||||||||||||||||
Db   1 WEAAAREACCRECCARA 17

Search completed: December 8, 2005, 16:22:17
Job time : 164 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 8, 2005, 15:49:22 ; Search time 46 Seconds
(without alignments)
30.554 Million cell updates/sec

Title: US-10-772-164-1
Perfect score: 101
Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	17	1	US-08-955-206-1	Sequence 1, Appli
2	101	100.0	17	2	US-08-955-050-1	Sequence 1, Appli
3	101	100.0	17	2	US-09-382-950-3	Sequence 3, Appli
4	101	100.0	17	2	US-09-382-736B-4	Sequence 4, Appli
5	101	100.0	17	2	US-09-406-781-48	Sequence 48, Appl
6	101	100.0	17	2	US-09-372-338-1	Sequence 1, Appli
7	101	100.0	17	2	US-09-880-132-48	Sequence 48, Appl
8	101	100.0	17	2	US-10-126-752-1	Sequence 1, Appli
9	101	100.0	17	2	US-09-502-664A-2	Sequence 2, Appli
10	101	100.0	17	2	US-09-813-197-4	Sequence 4, Appli
11	90	89.1	17	1	US-08-955-206-4	Sequence 4, Appli

12	90	89.1	17	2	US-08-955-050-4	Sequence 4, Appli
13	90	89.1	17	2	US-09-406-781-49	Sequence 49, Appl
14	90	89.1	17	2	US-09-372-338-4	Sequence 4, Appli
15	90	89.1	17	2	US-09-880-132-49	Sequence 49, Appl
16	90	89.1	17	2	US-10-126-752-4	Sequence 4, Appli
17	87	86.1	19	2	US-09-818-875-4368	Sequence 4368, Ap
18	56.5	55.9	106	2	US-09-252-991A-24846	Sequence 24846, A
19	54	53.5	245	2	US-09-270-767-35096	Sequence 35096, A
20	54	53.5	245	2	US-09-270-767-50313	Sequence 50313, A
21	53	52.5	365	2	US-09-252-991A-31971	Sequence 31971, A
22	52	51.5	631	2	US-09-252-991A-20063	Sequence 20063, A
23	50	49.5	28	2	US-08-486-099-161	Sequence 161, App
24	50	49.5	28	2	US-08-484-223B-161	Sequence 161, App
25	50	49.5	28	2	US-08-919-597-161	Sequence 161, App
26	50	49.5	28	2	US-08-475-668A-161	Sequence 161, App
27	50	49.5	28	2	US-08-485-551A-161	Sequence 161, App
28	50	49.5	28	2	US-08-471-913A-161	Sequence 161, App
29	50	49.5	28	2	US-08-485-264A-161	Sequence 161, App
30	50	49.5	28	2	US-09-082-279B-231	Sequence 231, App
31	50	49.5	28	2	US-08-474-349A-161	Sequence 161, App
32	50	49.5	28	2	US-09-315-304B-231	Sequence 231, App
33	50	49.5	28	2	US-08-973-952-14	Sequence 14, Appl
34	50	49.5	28	2	US-08-470-896-161	Sequence 161, App
35	50	49.5	28	2	US-08-485-546A-161	Sequence 161, App
36	50	49.5	28	2	US-09-834-784-231	Sequence 231, App
37	50	49.5	28	2	US-09-515-965A-231	Sequence 231, App
38	50	49.5	28	2	US-09-350-641C-231	Sequence 231, App
39	50	49.5	28	2	US-09-350-841A-231	Sequence 231, App
40	50	49.5	28	2	US-08-487-266A-161	Sequence 161, App
41	50	49.5	28	2	US-10-252-136-14	Sequence 14, Appl
42	50	49.5	28	2	US-08-484-741-161	Sequence 161, App
43	50	49.5	62	2	US-09-252-991A-28943	Sequence 28943, A
44	50	49.5	1380	2	US-09-949-016-11688	Sequence 11688, A
45	49.5	49.0	161	2	US-09-252-991A-28201	Sequence 28201, A
46	49	48.5	113	2	US-09-252-991A-19773	Sequence 19773, A
47	48	47.5	162	2	US-09-252-991A-30581	Sequence 30581, A
48	46	45.5	6	2	US-09-818-875-4385	Sequence 4385, Ap
49	46	45.5	101	2	US-09-199-637A-399	Sequence 399, App
50	46	45.5	129	2	US-09-252-991A-22496	Sequence 22496, A
51	46	45.5	147	2	US-09-252-991A-26082	Sequence 26082, A
52	46	45.5	227	2	US-09-252-991A-25546	Sequence 25546, A
53	46	45.5	498	4	PCT-US94-01101-2	Sequence 2, Appli
54	46	45.5	624	2	US-09-270-767-42659	Sequence 42659, A
55	46	45.5	1497	2	US-09-060-854B-2	Sequence 2, Appli
56	46	45.5	1497	2	US-09-529-904-3	Sequence 3, Appli
57	45	44.6	121	2	US-10-002-344A-257	Sequence 257, App
58	45	44.6	155	2	US-09-252-991A-30593	Sequence 30593, A
59	45	44.6	156	2	US-09-252-991A-21847	Sequence 21847, A
60	45	44.6	170	2	US-09-252-991A-20382	Sequence 20382, A
61	45	44.6	228	2	US-09-252-991A-30066	Sequence 30066, A
62	45	44.6	383	2	US-09-252-991A-29706	Sequence 29706, A
63	45	44.6	449	2	US-09-252-991A-23908	Sequence 23908, A
64	45	44.6	2088	2	US-09-548-372D-13	Sequence 13, Appl
65	45	44.6	2088	2	US-09-548-367D-13	Sequence 13, Appl
66	45	44.6	2088	2	US-09-551-853D-13	Sequence 13, Appl
67	45	44.6	2088	2	US-09-548-376D-13	Sequence 13, Appl
68	45	44.6	2088	2	US-09-548-373D-13	Sequence 13, Appl

69	45	44.6	2088	2	US-09-548-366F-13	Sequence 13, Appl
70	45	44.6	2088	2	US-09-548-368D-13	Sequence 13, Appl
71	44.5	44.1	99	2	US-09-270-767-44968	Sequence 44968, A
72	44.5	44.1	145	2	US-09-252-991A-29668	Sequence 29668, A
73	44	43.6	136	2	US-09-252-991A-23367	Sequence 23367, A
74	44	43.6	141	2	US-09-252-991A-20331	Sequence 20331, A
75	44	43.6	151	2	US-09-252-991A-32108	Sequence 32108, A
76	44	43.6	197	2	US-09-252-991A-32518	Sequence 32518, A
77	44	43.6	221	2	US-09-252-991A-21654	Sequence 21654, A
78	44	43.6	370	1	US-08-878-989-19	Sequence 19, Appl
79	44	43.6	370	2	US-09-272-796-19	Sequence 19, Appl
80	44	43.6	370	2	US-09-457-040B-31	Sequence 31, Appl
81	44	43.6	370	2	US-09-538-092-1314	Sequence 1314, Ap
82	44	43.6	415	2	US-09-949-016-7461	Sequence 7461, Ap
83	44	43.6	415	2	US-09-949-016-7462	Sequence 7462, Ap
84	44	43.6	485	2	US-10-214-269-20	Sequence 20, Appl
85	44	43.6	486	2	US-09-354-123-2	Sequence 2, Appli
86	44	43.6	1652	2	US-09-627-650B-1	Sequence 1, Appli
87	44	43.6	1652	2	US-09-436-063C-1	Sequence 1, Appli
88	44	43.6	1917	2	US-09-627-650B-5	Sequence 5, Appli
89	44	43.6	1917	2	US-09-436-063C-5	Sequence 5, Appli
90	44	43.6	2508	2	US-09-627-650B-7	Sequence 7, Appli
91	44	43.6	2508	2	US-09-436-063C-7	Sequence 7, Appli
92	44	43.6	2544	2	US-09-627-650B-3	Sequence 3, Appli
93	44	43.6	2544	2	US-09-436-063C-3	Sequence 3, Appli
94	44	43.6	2601	2	US-09-627-650B-9	Sequence 9, Appli
95	44	43.6	2601	2	US-09-436-063C-9	Sequence 9, Appli
96	43.5	43.1	218	2	US-09-252-991A-31933	Sequence 31933, A
97	43	42.6	21	2	US-09-488-799-73	Sequence 73, Appl
98	43	42.6	21	2	US-09-908-741-73	Sequence 73, Appl
99	43	42.6	26	2	US-09-073-407-13	Sequence 13, Appl
100	43	42.6	70	2	US-09-497-491-35	Sequence 35, Appl

ALIGNMENTS

```
RESULT 1
US-08-955-206-1
; Sequence 1, Application US/08955206
; Patent No. 5932474
;  GENERAL INFORMATION:
;    APPLICANT:  Tsien, Roger Y.
;    APPLICANT:  Griffin, B. Albert
;    TITLE OF INVENTION:  TARGET SEQUENCES FOR SYNTHETIC MOLECULES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
```

```
;       SOFTWARE:  FastSEQ for Windows Version 2.0b
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/955,206
;       FILING DATE:  21-OCT-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Haile, Ph.D., Lisa A.
;       REGISTRATION NUMBER:  38,347
;       REFERENCE/DOCKET NUMBER:  07257/060001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619/678-5070
;       TELEFAX:  619/678-5099
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  17 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FEATURE:
;       OTHER INFORMATION:  the N-terminus is acetylated and
;       OTHER INFORMATION:  the C-terminus is amidated
US-08-955-206-1

  Query Match             100.0%;  Score 101;  DB 1;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 9.2e-06;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 WEAAAREACCRECCARA 17
              |||||||||||||||||
Db          1 WEAAAREACCRECCARA 17



Search completed: December  8, 2005, 16:09:21
Job time : 48 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 8, 2005, 15:04:13 ; Search time 184 Seconds
(without alignments)
40.595 Million cell updates/sec

Title: US-10-772-164-1
Perfect score: 101
Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	17	2	AAY05336	Aay05336 Target se
2	101	100.0	17	3	AAB20847	Aab20847 Peptide a
3	101	100.0	17	4	AAB35430	Aab35430 Dye-bindi
4	101	100.0	17	4	AAM48100	Aam48100 Fluoresce
5	101	100.0	17	8	ADO06947	Ado06947 FLASH-bin
6	101	100.0	17	9	ADZ76895	Adz76895 RNA-tag f
7	90	89.1	17	2	AAY05337	Aay05337 Target se
8	90	89.1	17	3	AAB20848	Aab20848 Peptide a

9	87	86.1	19	4	AAM51838	Aam51838 Gene corr
10	87	86.1	19	5	AAU81286	Aau81286 Plasmid e
11	87	86.1	19	5	AAU75749	Aau75749 FLAsH pep
12	87	86.1	19	7	ADB78479	Adb78479 FIAsH pep
13	81	80.2	19	7	ABR84531	Abr84531 FLAsH pep
14	76	75.2	595	8	ADQ76865	Adq76865 Adenosine
15	61	60.4	22	3	AAY88739	Aay88739 Core poly
16	61	60.4	22	4	AAB77094	Aab77094 Core poly
17	61	60.4	22	4	ABB00098	Abb00098 Viral DP1
18	61	60.4	22	4	AAU12647	Aau12647 DP178-lik
19	61	60.4	55	5	ADE01583	Ade01583 Hybrid po
20	56.5	55.9	106	7	ABO76100	Abo76100 Pseudomon
21	53	52.5	365	7	ABO83225	Abo83225 Pseudomon
22	52	51.5	631	7	ABO71317	Abo71317 Pseudomon
23	51	50.5	535	8	ADL70535	Adl70535 Human G-p
24	50	49.5	28	3	AAY88872	Aay88872 Core poly
25	50	49.5	28	4	AAB77227	Aab77227 Core poly
26	50	49.5	28	4	ABB00231	Abb00231 Viral DP1
27	50	49.5	28	4	ABB01704	Abb01704 Viral cor
28	50	49.5	28	4	AAU12780	Aau12780 DP178-lik
29	50	49.5	28	6	ABO10317	Abo10317 HIV-1 BRU
30	50	49.5	30	8	ADT71522	Adt71522 Linker mo
31	50	49.5	32	8	ADT71523	Adt71523 Linker mo
32	50	49.5	35	8	ADT71524	Adt71524 Linker mo
33	50	49.5	62	7	ABO80197	Abo80197 Pseudomon
34	50	49.5	906	8	ADP31344	Adp31344 Human sec
35	50	49.5	1134	8	ADP30647	Adp30647 Human sec
36	49.5	49.0	161	7	ABO79455	Abo79455 Pseudomon
37	49	48.5	113	7	ABO71027	Abo71027 Pseudomon
38	49	48.5	120	2	AAW07542	Aaw07542 Clone 99,
39	49	48.5	918	8	ADP31459	Adp31459 Human sec
40	49	48.5	1626	8	ADP31008	Adp31008 Human sec
41	48	47.5	126	2	AAW98909	Aaw98909 Mouse IMC
42	48	47.5	131	2	AAW98908	Aaw98908 Mouse IMC
43	48	47.5	131	7	ADE25527	Ade25527 Mouse SLP
44	48	47.5	131	7	ADF28912	Adf28912 Mouse SLP
45	48	47.5	131	9	ADX02863	Adx02863 Murine an
46	48	47.5	146	8	ADQ59487	Adq59487 Human can
47	48	47.5	146	9	ADZ13856	Adz13856 Murine ca
48	48	47.5	162	7	ABO81835	Abo81835 Pseudomon
49	48	47.5	1305	8	ADP31389	Adp31389 Human sec
50	48	47.5	1312	8	ADP30999	Adp30999 Human sec
51	48	47.5	2001	8	ADP31644	Adp31644 Human sec
52	48	47.5	2260	8	ADP30687	Adp30687 Human sec
53	48	47.5	2272	8	ADP31136	Adp31136 Human sec
54	48	47.5	4440	6	ABU88256	Abu88256 Novel hum
55	48	47.5	4440	6	ABU90135	Abu90135 Novel hum
56	48	47.5	4440	6	ABU96437	Abu96437 Novel hum
57	48	47.5	4440	6	ABU99046	Abu99046 Novel hum
58	48	47.5	4440	6	ABU98261	Abu98261 Novel hum
59	48	47.5	4440	6	ABU91967	Abu91967 Novel hum
60	48	47.5	4440	6	ABU85271	Abu85271 Novel hum
61	48	47.5	4440	6	ABO00410	Abo00410 Novel hum
62	48	47.5	4440	6	ABU88961	Abu88961 Novel hum
63	48	47.5	4440	6	ABO06457	Abo06457 Novel hum
64	48	47.5	4440	6	ABU95517	Abu95517 Novel hum
65	48	47.5	4440	6	ABU95207	Abu95207 Novel hum

66	48	47.5	4440	6	ABU90755	Abu90755 Novel hum
67	48	47.5	4440	6	ABU93917	Abu93917 Novel hum
68	48	47.5	4440	6	ABU86191	Abu86191 Novel hum
69	48	47.5	4440	6	ABU82046	Abu82046 Novel hum
70	48	47.5	4440	6	ABU07907	Abu07907 Novel hum
71	48	47.5	4440	6	ABU94227	Abu94227 Novel hum
72	48	47.5	4440	6	ABO00100	Abo00100 Novel hum
73	48	47.5	4440	6	ABU87111	Abu87111 Novel hum
74	48	47.5	4440	6	ABU91352	Abu91352 Novel hum
75	48	47.5	4440	6	ABU90445	Abu90445 Novel hum
76	48	47.5	4440	6	ABU97036	Abu97036 Novel hum
77	48	47.5	4440	6	ABO05232	Abo05232 Novel hum
78	47	46.5	131	7	ADE25528	Ade25528 Rat SLPI
79	47	46.5	131	7	ADF28911	Adf28911 Rat SLPI
80	47	46.5	170	8	ADS10852	Ads10852 Human the
81	47	46.5	195	8	ADP30696	Adp30696 Human sec
82	47	46.5	205	8	ADS10854	Ads10854 Human the
83	47	46.5	210	9	AEA15447	Aea15447 Human pol
84	47	46.5	357	8	ADP31267	Adp31267 Human sec
85	47	46.5	621	8	ADP31147	Adp31147 Human sec
86	47	46.5	783	8	ADP31436	Adp31436 Human sec
87	47	46.5	821	8	ADP30679	Adp30679 Human sec
88	47	46.5	821	8	ADP30680	Adp30680 Human sec
89	47	46.5	882	8	ADP31688	Adp31688 Human sec
90	47	46.5	990	8	ADP31553	Adp31553 Human sec
91	47	46.5	1033	8	ADP30984	Adp30984 Human sec
92	47	46.5	1224	8	ADP31426	Adp31426 Human sec
93	47	46.5	1518	8	ADP31532	Adp31532 Human sec
94	47	46.5	1665	8	ADP31187	Adp31187 Human sec
95	47	46.5	1679	4	AAU07343	Aau07343 1-aminocy
96	47	46.5	2058	8	ADP31630	Adp31630 Human sec
97	47	46.5	2187	8	ADP30882	Adp30882 Human sec
98	47	46.5	3201	8	ADP31545	Adp31545 Human sec
99	47	46.5	3390	8	ADP31148	Adp31148 Human sec
100	47	46.5	3411	8	ADP30667	Adp30667 Human sec

ALIGNMENTS

RESULT 1
AAY05336
ID AAY05336 standard; peptide; 17 AA.
XX
AC AAY05336;
XX
DT 29-JUN-1999 (first entry)
XX
DE Target sequence peptide, SEQ ID NO. 1.
XX
KW Biarsenical compound; alpha-helix peptide; polypeptide purification;
KW immunoassay; crosslinking agent.
XX
OS Synthetic.
XX
PN WO9921013-A1.
XX

PD 29-APR-1999.
XX
PF 21-OCT-1998; 98WO-US022363.
XX
PR 21-OCT-1997; 97US-00955050.
PR 21-OCT-1997; 97US-00955206.
PR 21-OCT-1997; 97US-00955859.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Tsien RY, Griffin AB;
XX
DR WPI; 1999-288410/24.
XX
PT Biarsenical compounds that react specifically with cysteine residues.
XX
PS Claim 10; Page 41; 77pp; English.
XX
CC This sequence represents a target alpha-helix sequence for the
CC biarsenical compounds (BC) of the invention, which are able to react
CC specifically with cysteine residues in a target sequence to generate a
CC detectable signal. The BCs are used: (i) as labels that allow
CC identification of carrier molecules, e.g. in polypeptide purification,
CC immunoassays or other chemical or biological assays, including labelling
CC in vivo, e.g. to identify, locate or quantify polypeptides or nucleic
CC acids); (ii) for attaching a polypeptide to a solid substrate; or (iii)
CC to induce a polypeptide domain to adopt a more nearly alpha-helical form,
CC e.g. a conformation that can bind a drug. Tetra-arsenical compounds
CC derived from the BCs are used to crosslink two binding partners, e.g. to
CC study the effect of dimerisation on signal transduction. The BCs react
CC specifically with Cys-containing targets, and can be engineered to have
CC particular properties, especially ability to cross a biological membrane
CC and absence of any self-fluorescence. Both the BC and its target sequence
CC are small, and BC binding between them is reversible, e.g. by treatment
CC with a dithiol. Particularly, the BC becomes fluorescent when bound to
CC its target, but with a significant red-shift from the fluorescence of
CC fluorescein, allowing detection with very low background
XX
SQ Sequence 17 AA;

Query Match 100.0%; Score 101; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WEAAAREACCRECCARA 17
||||||||||||||||
Db 1 WEAAAREACCRECCARA 17

RESULT 2
AAB20847
ID AAB20847 standard; peptide; 17 AA.
XX
AC AAB20847;
XX
DT 03-JAN-2001 (first entry)
XX

DE Peptide amino acid sequence SEQ ID NO:48.
XX
KW Target protein binding element; protein level control; eukaryotic;
KW ubiquitination recognition element; treatment; infection; cancer;
KW inflammatory condition; genetic disease; insecticide; herbicide;
KW antiviral; antiparasitic; hepatotropic; antiinflammatory; cytostatic;
KW tumour; pest control; pesticide; rodenticide; fungicide; gene expression;
KW gene therapy.
XX
OS Unidentified.
XX
PN WO200047220-A1.
XX
PD 17-AUG-2000.
XX
PF 11-FEB-2000; 2000WO-US003436.
XX
PR 12-FEB-1999; 99US-0119851P.
PR 28-SEP-1999; 99US-00406781.
XX
PA (PROT-) PROTEINIX INC.
XX
PI Kenten JH, Roberts SF, Lebowitz MS;
XX
DR WPI; 2000-565258/52.
XX
PT Novel compounds for modulating the ubiquitination of target proteins
PT comprising a ubiquitination recognition element-target protein element
PT fusion, useful for treating viral infections.
XX
PS Disclosure; Page 55; 106pp; English.
XX
CC The present invention describes a compound (I) for activating the
CC ubiquitination (Ub'n) of a target protein comprising a Ub'n recognition
CC (peptide) element (URE) covalently linked to a target protein (peptide)
CC element (TPE). (I) can have antiviral, antiparasitic, hepatotropic,
CC antiinflammatory and cytostatic activities. The compound of (I) may be
CC used to treat a viral infection (especially hepatitis A, B, C or G, HIV-1
CC or 2, Herpes, CMV, rabies or Rouse sarcoma virus (RSV)), parasitic
CC infection, an infection caused by an eukaryotic organism in a mammal, to
CC treat a tumour or to control pests. The compound may also be used to
CC screen for target protein binding elements, to develop pesticides (e.g.
CC insecticides, rodenticides, fungicides and herbicides) and to control
CC gene expression (gene therapy). The present sequence represents an
CC example of a peptide which is given in the exemplification of the present
CC invention
XX
SQ Sequence 17 AA;

Query Match 100.0%; Score 101; DB 3; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WEAAAREACCRECCARA 17
||||||||||||||||
Db 1 WEAAAREACCRECCARA 17

RESULT 29
ABO10317
ID ABO10317 standard; peptide; 28 AA.
XX
AC ABO10317;
XX
DT 19-AUG-2003 (first entry)
XX
DE HIV-1 BRU gp41 DP178 based-peptide T234.
XX
KW HIV; DP107; DP178; glycoprotein 41; antiviral; virucide; EBV;
KW Epstein-Barr virus infection; heptad repeat motif.
XX
OS Human immunodeficiency virus; isolate BRU.
OS Synthetic.
XX
PN US6518013-B1.
XX
PD 11-FEB-2003.
XX
PF 07-JUN-1995; 95US-00485546.
XX
PR 07-JUN-1993; 93US-00073028.
PR 07-JUN-1994; 94US-00255208.
PR 20-DEC-1994; 94US-00360107.
XX
PA (TRIM-) TRIMERIS INC.
XX
PI Barney SO, Lambert DM, Petteway SR;
XX
DR WPI; 2003-465599/44.
XX
PT Inhibiting transmission of Epstein-Barr virus to a cell, by contacting
PT the cell with a peptide consisting of a region of Epstein-Barr virus
PT protein.
XX
PS Example; Fig 49G; 716pp; English.
XX
CC The invention relates to inhibiting (M) transmission of an Epstein-Barr
CC virus to a cell, comprising contacting the cell with an effective
CC concentration of a peptide consisting of a region of 16-39 consecutive
CC amino acids of an Epstein-Barr virus protein for an effective period of
CC time, where the region is recognised by one or more of ALLMOTI5,
CC 107x178x4 or PLZIP sequence search motifs, the peptide further comprises
CC an amino terminal X, and a carboxy terminal Z in which X comprises an
CC amino group, acetyl group, 9-fluorenylmethoxy-carbonyl group, hydrophobic
CC group or macromolecular carrier group, and Z comprises a carboxyl group,
CC amido group, hydrophobic group, or macromolecular carrier group, and
CC fusion of the virus to the cell is inhibited. The peptides were
CC identified by analysing the structure/motifs present in the HIV-1
CC glycoprotein 41 anti-HIV peptides DP107 and DP178. These heptad repeat
CC motif containing peptides were used to design the motifs cited above,
CC which in turn were used to analyse proteins from other pathogenic
CC organisms and HIV isolates, looking for DP107/178 structural analogues.
CC The method is useful for inhibiting transmission of Epstein-Barr virus to

CC a cell and Epstein-Barr virus infection. The present sequence is a
CC antiviral peptide based on a region of a protein from a pathogenic
CC organism analogous to DP107 or DP178
XX
SQ Sequence 28 AA;

Query Match 49.5%; Score 50; DB 6; Length 28;
Best Local Similarity 73.3%; Pred. No. 33;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 EAAAREACCRECCAR 16
||||||| || ||
Db 1 EAAAREAAAREAAAR 15

RESULT 38
AAW07542
ID AAW07542 standard; protein; 120 AA.
XX
AC AAW07542;
XX
DT 07-FEB-1997 (first entry)
XX
DE Clone 99, human pro-opiomelanocortin cDNA analogue protein prod. (2).
XX
KW Human; poly(A) RNA; cDNA synthesis; polymerase chain reaction;
KW lambda gt11; phage vector; PCR; amplification; clone 99;
KW pro-opiomelanocortin.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 101
FT /note= "corresponding codon TGA"
XX
PN EP716150-A1.
XX
PD 12-JUN-1996.
XX
PF 05-DEC-1995; 95EP-00119121.
XX
PR 05-DEC-1994; 94JP-00300657.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Onda H, Hosoya M;
XX
DR WPI; 1996-269991/28.
DR N-PSDB; AAT43979.
XX
PT DNA primers for sequences encoding Gly-Lys-Arg, Gly-Arg-Arg or Gly-Lys-
PT Lys - useful for identifying peptide(s) with useful physiological
PT activity having the specified sequences at their C-terminal ends.
XX
PS Example 3; Fig 10; 37pp; English.
XX
CC Human poly(A) RNA was used as a template for cDNA synthesis, conducted by
CC using, as primers, antisense codons for Gly-Lys-Arg, Gly-Arg-Arg or Gly-

CC Lys-Lys. The prod. was ligated into a lambda gt11 phage vector, and PCR
CC amplified. The prod. was subcloned with a TA receptor, and cDNA fragments
CC from 100 clones sequenced, including clone 99, which was decoded in 3
CC reading frames to give AAW07541-43. The nucleotide sequence of clone 99
CC was found to have a portion of cDNA encoding human pro-opiomelanocortin,
CC an entire sequence of cDNA encoding gamma-MSH and a sequence identical
CC with the 5'-upstream region of Gly-Arg-Arg
XX
SQ Sequence 120 AA;

Query Match 48.5%; Score 49; DB 2; Length 120;
Best Local Similarity 64.3%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy          3 AAAREACCRECCAR 16
              ||||  ||  || |
Db         37 AAARGPCCWPCCFR 50
```

Search completed: December 8, 2005, 16:03:54
Job time : 189 secs